



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 118756**

**TO: Michael Borin**  
**Location: rem/2a55/2c70**  
**Art Unit: 1631**  
**Tuesday, April 06, 2004**

**Case Serial Number: 09/997807**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**REM-1A55**  
**Phone: 571-272-2512**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Borin,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

• Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

***Because they contain data that is confidential, the results of Pending database searches should not be left in the case .***

STIC-Biotech/ChemLib

118756

From: Borin, Michael  
Sent: Monday, April 05, 2004 6:04 PM  
To: STIC-Biotech/ChemLib  
Subject: Search request:09/997807

Examiner: M.Borin  
AU: 1631

Remsen 2A55  
Tel.: 20713

RE: 09/997807; peptide polymer

Please conduct search of polypeptide SEQ ID No. 2 against the commercial and interference protein databases.

Thank you

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 4/5/04  
Date Completed: 4/6/04  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: osp  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
571-272-2507 Remsen E01 D86

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:24:31 ; Search time 60 Seconds  
(without alignments)  
974.789 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VXYTTLAIAGIIASAALAL.....EGMLFDSLFIIVFQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	1042	100.0	207	5	AAU99735	Pyrodicti
2	1042	100.0	207	7	ADB46044	CanA fusi
3	553.5	53.1	170	5	AAU99736	Pyrodicti
4	553.5	53.1	170	7	ADB46046	CanB fusi
5	540	51.8	178	5	AAU99737	Pyrodicti
6	540	51.8	178	7	ADB46048	CanC fusi
7	305.5	29.3	124	5	AAU99739	Pyrodicti
8	305.5	29.3	124	7	ADB46052	CanE fusi
9	304.5	29.2	130	5	AAU99738	Pyrodicti
10	304.5	29.2	130	7	ADB46050	CanD fusi
11	96	9.2	255	7	ADE45056	CAMP fact
12	92	8.8	322	6	ABU33304	Protein e
13	91	8.7	255	5	ABP30785	Streptoco
14	87.5	8.4	978	7	ADC51492	Bacterial
15	86.5	8.3	284	5	ABBS7211	Mouse isc
16	86.5	8.3	418	3	AAU77286	Streptomy
17	86.5	8.3	418	3	AAU78838	Amino aci
18	86.5	8.3	846	5	AAU82974	S. cerevi
19	86.5	8.3	846	6	ABR53317	Protein s
20	85.5	8.2	416	5	ABP65732	Bifidobac
21	85	8.2	338	5	ABR93924	Herbicida
22	83	8.0	626	2	AAU98010	BCG Ag85B
23	83	8.0	626	6	ABU08321	M. bovis
24	83	8.0	675	5	ABP65592	Bifidobac
25	83	8.0	1246	5	AAU84295	Human end

## ALIGNMENTS

### RESULT 1

AAU99735  
ID AAU99735 standard; protein; 207 AA.

XX AC AAU99735;

XX AC (first entry)

DT 07-OCT-2002 (first entry)

XX Pyrodicticum abyssi Canulle A (CanA) protein.

XX Polymer; self-assembly; divalent cation; drug delivery; human body;  
XX animal body; fibre; polymeric separation agent; coating composition;  
XX biochip; nanomechanical component; optical switch; optical wave guide;  
XX Canulle A; CanA.

OS Pyrodicticum abyssi.

XX WO200244336-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US045001.

XX 30-NOV-2000; 2000US-0250426P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EJ, Laiferty WM, Barton N, Chow K;

XX WPI; 2002-557536/53.

XX N-PSDB; ABK88271.

XX Producing a polypeptide polymer by self-assembly for use in lubricants  
XX and coating compositions, comprises polymerizing polypeptides capable of  
XX self-assembly in the presence of a divalent cation and template molecule.

XX Claim 9; Page 178-179; 182pp; English.

XX The present invention relates to a new method of producing a polypeptide  
XX polymer by self-assembly. The method involves providing a number of  
XX polypeptides capable of self-assembly in the presence of a divalent  
XX cation and polymerising the polypeptides in the presence of a divalent  
XX cation and a template molecule. The invention is useful for delivering a  
XX drug to a location in the human or animal body. Polypeptides are useful  
XX for encapsulating a molecule. The polymeric separation agent is useful  
XX for isolating a chiral compound from a mixture. A nucleic acid is useful  
XX for comparing a first sequence to a second sequence, where the first  
XX sequence is a nucleic acid, and for identifying a feature in a particular

Abb30742 Human Tum  
Aaol7366 Human nid  
Abus4449 Human tum  
Abp27517 Streptoco  
Aag51600 Arabidops  
Aag06413 Arabidops  
Aag51599 Arabidops  
Aag06412 Arabidops  
Aae36342 Arabidops  
Abu24247 Protein e  
Aaw43448 Tobacco 1  
Abb47321 Listeria  
Abb47302 Listeria  
Abp73966 Candida a  
Abus0834 Helicobac  
Aa051217 Helicobac  
Aar05710 TRY40. 3/  
Aar06478 TRY40. 3/  
Aar43675 Single ch  
Aar99645 Single ch

26 83 1247 5 ABB90742  
27 83 1247 5 AAO17366  
28 83 1247 6 ABUS4449  
29 83 1370 5 ABP27517  
30 82.5 7.9 238 3 AAG51600  
31 82.5 7.9 238 3 AAG06413  
32 82.5 7.9 243 3 AAG51599  
33 82.5 7.9 243 3 AAG06412  
34 82.5 7.9 243 6 AAE36342  
35 82.5 7.9 296 6 ABU24247  
36 81.5 7.8 557 2 AAW43448  
37 81.5 7.8 862 5 ABB47321  
38 81.5 7.8 2044 5 ABB47302  
39 81 7.8 751 5 ABP73966  
40 80.5 7.7 148 5 ABUS0834  
41 80.5 7.7 183 5 ABUS1217  
42 80.5 7.7 225 2 AAR05710  
43 80.5 7.7 225 2 AAR06478  
44 80.5 7.7 225 2 AAR43675  
45 80.5 7.7 225 2 AAR99645

CC sequence. The polypeptide of the invention is useful in fibres, polymeric  
 CC separation agents, coating compositions, biochips, nanomechanical  
 CC components, optical switches and optical wave guides. The present amino  
 CC acid sequence represents the Pyrodictium abyssi Cannule A (CanA) protein  
 CC of the invention  
 XX  
 XX SQ Sequence 207 AA;

Query Match 100.0%; Score 1042; DB 5; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-95;  
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTTIAAGIIASAAALALLAGFATTQSPINSFYATGTAAQVSEPIDVESHLSITPAA 60  
 DB 1 VKYTTIAAGIIASAAALALLAGFATTQSPINSFYATGTAAQVSEPIDVESHLSITPAA 60  
 QY 61 GAQSDDIGVAIVWIKQVNDVKLVLRNAEQLKPYFKYLOIQTSGYETNSTALGNFS 120  
 DB 61 GAQSDDIGVAIVWIKQVNDVKLVLRNAEQLKPYFKYLOIQTSGYETNSTALGNFS 120  
 QY 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTGYTNTSIWVGPEDKIIYVNETKPVAILNF 180  
 DB 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTGYTNTSIWVGPEDKIIYVNETKPVAILNF 180  
 QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207  
 DB 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207

RESULT 2  
 ADB46044  
 ID : ADB46044 standard; protein; 207 AA.

XX AC ADB46044;  
 XX  
 XX  
 DT 04-DEC-2003 (first entry)  
 DE CanA fusion protein.  
 XX  
 XX  
 KW chimeric cannulae polypeptide; metabolite exchange; signal compound;  
 KW CanA.

OS Pyrodictium abyssi.  
 XX  
 XX WO2003070961-A1.  
 XX  
 XX 28-AUG-2003.  
 XX  
 XX 18-FEB-2003; 2003WO-US004828.  
 XX  
 XX 15-FEB-2002; 2002US-0357406P.  
 XX  
 XX (DIVE-) DIVERSA CORP.  
 XX  
 PI Barton NR, O'donoghue E, Frey G;  
 XX  
 XX WPI; 2003-689787/65.  
 DR N-PSDB; ADB46043.  
 XX

PT New chimeric polypeptide comprising a first domain comprising a cannulae  
 PT polypeptide and a second domain comprising a heterologous polypeptide or  
 PT peptide useful for selecting and purifying chiral compositions from  
 PT racemic mixtures.

XX  
 PS Claim 3; Page 81-82; 89pp; English.  
 CC  
 CC The present invention relates to a chimeric cannulae polypeptide and a  
 CC second domain having a heterologous polypeptide or peptide. The  
 CC polypeptides, nanotubule and methods are useful for identifying,  
 CC separating and synthesizing proteins or ligands, and for selecting  
 CC purifying chiral compositions from racemic mixtures. The nanotubule is  
 CC useful for cells to exchange metabolites, genetic information or signal  
 CC compounds. The present sequence represents CanA fusion protein.

XX SQ Sequence 207 AA;

Query Match 100.0%; Score 1042; DB 7; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-95;  
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTTIAAGIIASAAALALLAGFATTQSPINSFYATGTAAQVSEPIDVESHLSITPAA 60  
 DB 1 VKYTTIAAGIIASAAALALLAGFATTQSPINSFYATGTAAQVSEPIDVESHLSITPAA 60  
 QY 61 GAQSDDIGVAIVWIKQVNDVKLVLRNAEQLKPYFKYLOIQTSGYETNSTALGNFS 120  
 DB 61 GAQSDDIGVAIVWIKQVNDVKLVLRNAEQLKPYFKYLOIQTSGYETNSTALGNFS 120  
 QY 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTGYTNTSIWVGPEDKIIYVNETKPVAILNF 180  
 DB 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTGYTNTSIWVGPEDKIIYVNETKPVAILNF 180  
 QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207  
 DB 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207

RESULT 3  
 AAU99736  
 ID : AAU99736 standard; protein; 170 AA.

XX AC AAU99736;  
 XX  
 XX  
 DT 07-OCT-2002 (first entry)  
 DE Pyrodictium abyssi Cannule B (CanB) protein.

XX Polymer; self-assembly; divalent cation; drug delivery; human body;  
 KW animal body; fibre; polymeric separation agent; coating composition;  
 KW biochip; nanomechanical component; optical switch; optical wave guide;  
 KW Cannule B; CanB.

XX Pyrodictium abyssi.

XX WO200244336-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US045001.

XX 30-NOV-2000; 2000US-0250426P.

XX (DIVE-) DIVERSA CORP.

XX Short J, Mathur EJ, Lafferty WM, Barton N, Chow K;

XX WPI; 2002-557536/59.

XX N-PSDB; ABK88272.

XX Producing a polypeptide polymer by self-assembly for use in lubricants  
 XX and coating compositions, comprises polymerizing polypeptides capable of  
 XX self-assembly in the presence of a divalent cation and template molecule.

XX Claim 9; Page 179-180; 182pp; English.

XX The present invention relates to a new method of producing a polypeptide  
 XX polymer by self-assembly. The method involves providing a number of  
 XX polypeptides capable of self-assembly in the presence of a divalent  
 XX cation and polymerising the polypeptides in the presence of a divalent  
 XX cation and a template molecule. The invention is useful for delivering a  
 XX drug to a location in the human or animal body. Polypeptides are useful  
 XX for encapsulating a molecule. The polymeric separation agent is useful  
 XX for isolating a chiral compound from a mixture. A nucleic acid is useful  
 XX for comparing a first sequence to a second sequence, where the first  
 XX sequence is a nucleic acid, and for identifying a feature in a particular  
 XX sequence. The polypeptide of the invention is useful in fibres, polymeric

CC separation agents, coating compositions, biochips, nanomechanical  
 CC components, optical switches and optical wave guides. The present amino  
 CC acid sequence represents the Pyrodictium abyssii Cannule B (CanB) protein  
 CC of the invention

XX SQ Sequence 170 AA;  
 Query Match 53.1%; Score 553.5; DB 5; Length 170;  
 Best Local Similarity 60.2%; Pred. No. 1.7e-46;  
 Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;

Qy 1 VKYTLAIGIISAAALALLAGFATTQSPPLNSFYATGTAQAVSEPIDVSHLSITPAA 60  
 Db 1 VKPTALALAGIISAADLALLAGFATTQSPPLNSFYATGTAQAVSEPIDVSHLSITPAA 60

Qy 61 GAQSDDIGYAIWIKQVNDVKLVLRNAEQKPYFKYLOIQITSGYETNSTALGNFS 120  
 Db 61 GAQSQSDIGYFNVTAKDQVNVTKIKVTLANAQKPYFKYLOIQLKSEVA-----D 111

Qy 121 ETKAVISLDNPSAVIVLDKEDIAVLDPDKTYNTSIWVGPEDPKIIVYNETKPVAILNF 180  
 Db 112 EIKAVISIDKPSAVIILDSQDF-----DSNNRAKISA 143

Qy 181 KAFYEAKGMLFDSLPIFNQVLOV 206  
 Db 144 TAYEAKGMLFDSLPIFNQVLSV 169

RESULT 4  
 ADB46046  
 ID ADB46046 standard; protein; 170 AA.  
 AC ADB46046;  
 XX AC  
 DT 04-DEC-2003 (first entry)  
 XX DE  
 DE CanB fusion protein.  
 XX KW  
 KW chimeric cannulae polypeptide; metabolite exchange; signal compound;  
 KW CanB.  
 XX OS  
 OS Pyrodictium abyssii.  
 XX PN  
 PN WO2003070961-A1.  
 XX PD  
 PD 28-AUG-2003.  
 XX PF  
 PF 18-FEB-2003; 2003WO-US004828.  
 XX PR  
 PR 15-FEB-2002; 2002US-0357406P.  
 XX PA  
 PA (DIVE-) DIVERSA CORP.  
 XX PI  
 PI Barton NR, O'donoghue E, Frey G;  
 XX DR  
 DR WPI; 2003-689787/65.  
 XX DR  
 DR N-PSDB; ADB46045.  
 XX PT  
 PT New chimeric polypeptide comprising a first domain comprising a cannulae  
 PT polypeptide and a second domain comprising a heterologous polypeptide  
 PT peptide useful for selecting and purifying chiral compositions from  
 PT racemic mixtures.  
 XX PS  
 PS Claim 3; Page 82-83; 89pp; English.  
 XX CC  
 CC The present invention relates to a chimeric cannulae polypeptide and a  
 CC second domain having a heterologous polypeptide or peptide. The  
 CC polypeptides, nanotubule and methods are useful for identifying,  
 CC separating and synthesizing proteins or ligands, and for selecting and  
 CC purifying chiral compositions from racemic mixtures. The nanotubule is  
 CC useful for cells to exchange metabolites, genetic information or signal  
 CC compounds. The present sequence represents CanB fusion protein.

XX SQ Sequence 170 AA;  
 Query Match 53.1%; Score 553.5; DB 7; Length 170;  
 Best Local Similarity 60.2%; Pred. No. 1.7e-46;  
 Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;

Qy 1 VKYTLAIGIISAAALALLAGFATTQSPPLNSFYATGTAQAVSEPIDVSHLSITPAA 60  
 Db 1 VKPTALALAGIISAADLALLAGFATTQSPPLNSFYATGTAQAVSEPIDVSHLSITPAA 60

Qy 61 GAQSDDIGYAIWIKQVNDVKLVLRNAEQKPYFKYLOIQITSGYETNSTALGNFS 120  
 Db 61 GAQSQSDIGYFNVTAKDQVNVTKIKVTLANAQKPYFKYLOIQLKSEVA-----D 111

Qy 121 ETKAVISLDNPSAVIVLDKEDIAVLDPDKTYNTSIWVGPEDPKIIVYNETKPVAILNF 180  
 Db 112 EIKAVISIDKPSAVIILDSQDF-----DSNNRAKISA 143

Qy 181 KAFYEAKGMLFDSLPIFNQVLOV 206  
 Db 144 TAYEAKGMLFDSLPIFNQVLSV 169

RESULT 5  
 AAU99737  
 ID AAU99737 standard; protein; 178 AA.  
 AC AAU99737;  
 XX AC  
 DT 07-OCT-2002 (first entry)  
 XX DE  
 DE Pyrodictium abyssii Cannule C (CanC) protein.  
 XX KW  
 KW Polymer; self-assembly; divalent cation; drug delivery; human body;  
 KW animal body; fibre; polymeric separation agent; coating composition;  
 KW biochip; nanomechanical component; optical switch; optical wave guide;  
 KW Cannule C; CanC.  
 XX OS  
 OS Pyrodictium abyssii.  
 XX PN  
 PN WO200244336-A2.  
 XX PD  
 PD 06-JUN-2002.  
 XX PF  
 PF 30-NOV-2001; 2001WO-US045001.  
 XX PR  
 PR 30-NOV-2000; 2000US-0250426P.  
 XX PA  
 PA (DIVE-) DIVERSA CORP.  
 XX PI  
 PI Short J, Mathur EJ, Lafferty WM, Barton N, Chow K;  
 XX DR  
 DR WPI; 2002-557536/59.  
 XX DR  
 DR N-PSDB; ABK88273.  
 XX PT  
 PT Producing a polypeptide polymer by self-assembly for use in lubricants  
 PT and coating compositions, comprises polymerizing polypeptides capable of  
 PT self-assembly in the presence of a divalent cation and template molecule.  
 XX PS  
 PS Claim 9; Page 180; 182pp; English.  
 XX CC  
 CC The present invention relates to a new method of producing a polypeptide  
 CC polymer by self-assembly. The method involves providing a number of  
 CC polypeptides capable of self-assembly in the presence of a divalent  
 CC cation and polymerising the polypeptides in the presence of a divalent  
 CC cation and a template molecule. The invention is useful for delivering a  
 CC drug to a location in the human or animal body. Polypeptides are useful  
 CC for encapsulating a molecule. The polymeric separation agent is useful  
 CC for isolating a chiral compound from a mixture. A nucleic acid is useful  
 CC for comparing a first sequence to a second sequence, where the first  
 CC sequence is a nucleic acid, and for identifying a feature in a particular  
 CC sequence. The polypeptide of the invention is useful in fibres, polymeric  
 CC separation agents, coating compositions, biochips, nanomechanical

CC components, optical switches and optical wave guides. The present amino  
CC acid sequence represents the Pyrodictium abyssi Cannule C (CanC) protein  
CC of the invention

Sequence 178 AA;

[illegible]

RESULT 6  
ADB46048  
ID ADB46048 standard; protein; 178 AA.  
XX  
XX ADB46048;  
XX  
XX  
DT 04-DEC-2003 (first entry)  
DT  
XX  
XX  
DE CanC fusion protein.

XX  
DR WFI; 2003-689787/65.  
DR N-PSDB; ADB46047.  
XX  
PT New chimeric polypeptide comprising a first domain comprising a cannulae  
PT polypeptide and a second domain comprising a heterologous polypeptide or  
PT peptide useful for selecting and purifying chiral compositions from  
PT racemic mixtures.

XX  
PS  
Claim 3; Page 83-84; 89pp; English.

The present invention relates to a chimeric cannulae polypeptide and a second domain having a heterologous polypeptide or peptide. The polypeptides, nanotubule and methods are useful for identifying, separating and synthesizing proteins or ligands, and for selecting and purifying chiral compositions from racemic mixtures. The nanotubule is useful for cells to exchange metabolites, genetic information or signal compounds. The present sequence represents a Cdc fusion protein.

Sequence 178 AA;  
SQ

```

Query Match          51.8%; Score 540; DB 7; Length 178;
Best Local Similarity 58.5%; Pred.No.3.9e-45;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY      1 VKYTTLAIAGIIASAALALLACGATTQSPINSEFYATGTAAVSEPIDVESHL-GSITPA 59
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      1 MRYTTLAIGIVASAAALALLACGATTQSPISFAYATGTAAVSEPIDVESHLDTIAPA 60
       :|::||::||::||::||::||::||::||::||::||::||::||::||:

QY     60 AGAQGGDDITGYALVIWKIQDQNVKLVKLRNAQLKPYPKYLIQITSGYETNSTALGNF 119
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db     61 AGAQGYKDMGYIKITNQSKVNVIKLVLANEAQLKPFPDYQLQVLTS-----NATGT--- 113
       :|::||::||::||::||::||::||::||::||::||::||::||::||:

QY    120 SEYKAVISLDNRSAVILDKEDIAVLVPDKTYNTSIWVGPEPDKTIIVNETKPVAILN 179
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    114 DMVKAVLSLEKPSAVILNDND-----YDSTNKIQ-LK 145
       :|::||::||::||::||::||::||::||::||::||::||::||::||:

QY    180 FKAFYEAKEGMFLFDSLPVFIFNFQVL 204
       :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db    146 VEAYFEAKEGMFLFDSLPIVILNFQVL 170
       :|::||::||::||::||::||::||::||::||::||::||::||::||:

```

RESULT 7  
AAU99739  
ID AAU99739 standard; protein: 124 AA.

07-OCT-2002	(first entry)
Pyrodicticum abyssi	Cannule E (CanE) partial protein.
Polymer;	self-assembly; divalent cation; drug delivery
animal body;	fibre; polymeric separation agent; coating
biochip;	nanomechanical component; optical switch; optical
Cannule E;	CanE partial.

AA  
PN  
WO200244336-A2.

AA 06-JUN-2002.

30-NOV-2001; 2001WO-US045001.

PR 30-NOV-2000; 2000US-0250426P.

PA (DIVE-) DIVERSA CORP.

PI Short J, Mathur EJ, Lafferty WM, Barton N, Chow K;

DR WPI; 2002-557536/59.

Producing a polypeptide polymer by self-assembly for use in lubricants and coating compositions, comprises polymerizing polypeptides capable of self-assembly in the presence of a divalent cation and template molecule

[illegible]

The present invention relates to a new method of producing a polypeptide polymer by self-assembly. The method involves providing a number of polypeptides capable of self-assembly in the presence of a divalent cation and polymerising the polypeptides in the presence of a divalent cation and a template molecule. The invention is useful for delivering a drug to a location in the human or animal body. Polypeptides are useful for encapsulating a molecule. The polymeric separation agent is useful for isolating a chiral compound from a mixture. A nucleic acid is useful for comparing a first sequence to a second sequence, where the first sequence is a nucleic acid, and for identifying a feature in a particular sequence. The polypeptide of the invention is useful in fibres, polymeric separation agents, coating compositions, biochips, nanomechanical components, optical switches and optical wave guides. The present amino



[illegible]

Best Local Similarity 47.2%; Pred. No. 5.1e-22;  
Matches 75; Conservative 22; Mismatches 33; Indels 29; Gaps 6;

QY 33 SFYATGTAQAVSEPIDVESHGISTPAAGAGSDDIGVAIVWIKQDVNDVKLVKTLNAAE 92  
DB 1 SFYATGTAQAVSEPIDVSSIGTLNTAAGAGKQTLGDITIIYAHNDVNTIKLVTLANAA 60

QY 93 QLKPYFKYLOITSGVYETNSALGNFSETKAVISLDNPSAVIVLDKEDIAVLVLPDKTGY 152  
DB 61 QLRPFYKYLIIKLVS-LDSN---GNESEKGMITLWPKYAVIILDHED-----F 105

QY 153 TNSIWIWPGEPDKIIVNETKPVAILNFKAFYEAKGML 191  
DB 106 NND---IDGD-----NQCQ-----IDATAYYEAKGML 130

RESULT 10  
ID ADB46050 standard; protein; 130 AA.  
XX ADB46050;  
AC  
XX  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE CanD fusion protein.  
XX  
XX chimeric cannulae polypeptide; metabolite exchange; signal compound;  
XX CanD.  
OS Pyrodictium abyssi.  
XX  
XX WO2003070961-A1.  
XX  
XX  
XX 28-AUG-2003.  
XX  
XX 18-FEB-2003; 2003WO-US004828.  
XX  
XX 15-FEB-2002; 2002US-0357406P.  
XX  
XX (DIVE-) DIVERSA CORP.  
XX  
XX Barton NR, O'donoghue E, Frey G;  
XX WPI; 2003-689787/65.  
XX N-PSDB; ADB46049.  
XX  
XX New chimeric polypeptide comprising a first domain comprising a cannulae polypeptide and a second domain comprising a heterologous polypeptide or peptide useful for selecting and purifying chiral compositions from racemic mixtures.  
XX  
XX Disclosure; Page 84; 89pp; English.  
XX  
XX The present invention relates to a chimeric cannulae polypeptide and a second domain having a heterologous polypeptide or peptide. The polypeptides, nanotubule and methods are useful for identifying, separating and synthesizing proteins or ligands, and for selecting and purifying chiral compositions from racemic mixtures. The nanotubule is useful for cells to exchange metabolites, genetic information or signal compounds. The present sequence represents CanD fusion protein.  
XX  
XX Sequence 130 AA;  
PS  
PS Query Match 29.2%; Score 304.5; DB 7; Length 130;  
Best Local Similarity 47.2%; Pred. No. 5.1e-22;  
Matches 75; Conservative 22; Mismatches 33; Indels 29; Gaps 6;

QY 33 SFYATGTAQAVSEPIDVESHGISTPAAGAGSDDIGVAIVWIKQDVNDVKLVKTLNAAE 92  
DB 1 SFYATGTAQAVSEPIDVSSIGTLNTAAGAGKQTLGDITIIYAHNDVNTIKLVTLANAA 60

QY 93 QLKPYFKYLOITSGVYETNSALGNFSETKAVISLDNPSAVIVLDKEDIAVLVLPDKTGY 152  
DB 61 QLRPFYKYLIIKLVS-LDSN---GNESEKGMITLWPKYAVIILDHED-----F 105

QY 153 TNSIWIWPGEPDKIIVNETKPVAILNFKAFYEAKGML 191  
DB 106 NND---IDGD-----NQCQ-----IDATAYYEAKGML 130

RESULT 11  
ID ADE45056 standard; protein; 255 AA.  
XX ADE45056;  
AC  
XX  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE CAMP factor, SEQ ID 4.  
XX  
XX Immunogenic; cyclic adenosine monophosphate factor; CAMP factor; mastitis; listeriosis; antibacterial; antiinflammatory; vaccine; immunostimulant; gene therapy.  
XX  
XX Streptococcus agalactiae.  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..29  
FT /note= "Signal peptide"  
FT Protein 30..255  
FT /note= "Mature protein"  
XX  
XX WO2003091437-A1.  
XX  
XX 06-NOV-2003.  
XX  
XX 24-APR-2003; 2003WO-CA000587.  
XX  
XX 26-APR-2002; 2002US-00134021.  
XX  
XX (UYSA-) UNIV SASKATCHEWAN.  
XX  
XX Potter AA, Perez-Casal J, Fontaine M, Song X;  
XX WPI; 2003-877332/81.  
XX N-PSDB; ADE45055.  
XX  
XX New cyclic adenosine monophosphate (CAMP)-3 polypeptide comprising streptococcal CAMP factor epitopes, useful in treating or preventing bacterial infection or as vaccine or diagnostic reagent.  
XX  
XX Claim 7; Fig 3; 73pp; English.  
XX  
XX The present invention relates to immunogenic proteins comprising one or more cyclic adenosine monophosphate (CAMP) factor epitopes from more than one bacterial species. The proteins and CAMP factors are useful in detecting Streptococcus antibodies in a biological sample, eliciting an immunological response against streptococcal infections, used in vaccine compositions or as diagnostic reagents or in preventing or treating streptococcal infection which causes mastitis. The proteins are also useful in preventing or treating listeriosis. The present sequence is a CAMP factor.  
XX  
XX Sequence 255 AA;  
PS  
PS Query Match 9.2%; Score 96; DB 7; Length 255;  
Best Local Similarity 25.0%; Pred. No. 0.59;  
Matches 53; Conservative 31; Mismatches 84; Indels 44; Gaps 10;

QY 5 TLAAGIIASAALALAGFATTQSPNSFYATGTAAVSEPIDVES-HLGSITPAAGAQ 63  
DB 12 TLVAGALLSPAVLEHDAQVTTTPQVNVHNSNQQAQKLQDSIQLRNKKD--NVQ 69

QY 64 GSD---DIGVAIVWIKQDVNDVKLVKTLR-----NARQLKPYFKYLOIT 106  
DB 70 GTDYKPVNEAITSVE-----KLKTSLRANPETVYDLNSIGSRVEALTDVIE--AITFS 121

QY 107 SGVETNSTALGN-----FSETKAVISLONPSAVI-----VLDKEDIADVLPDKTGYT 153  
 Db 122 TQHLTNTKVSQANIDMGFGITKLVRIDDPFASVDSIKAQVNDVKALQKVLTPDLKPTD 181  
 QY 154 NTSIWPGEPPDKIIVYNE--TKPVAILNFKAF 183  
 Db 182 RATIYTKSLDKET-WNTRFTRDKKVLNVKEF 212

## RESULT 12

ABU33304  
 ID ABU33304 standard; protein; 322 AA.

XX AC ABU33304;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #18831.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Legionella pneumophila.

XX PN WO200277183-A2.

XX PD 03-OCT-2002;

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA37174.

XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 XX PT for homologous nucleic acids required for cellular proliferation to  
 XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 61228; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of  
 XX CC the 6213 antisense sequences given in the specification where expression  
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 XX CC encoding a polypeptide whose expression is inhibited by the antisense  
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 XX CC polypeptide or its fragment whose expression is inhibited by the  
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 XX CC proliferation or the activity of a gene in an operon required for  
 XX CC proliferation; (7) identifying a compound that influences the activity of  
 XX CC the gene product or that has an activity against a biological pathway  
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
 XX CC identifying a gene required for cellular proliferation or the biological  
 XX CC pathway in which a proliferation-required gene or its gene product lies  
 XX CC or a gene on which the test compound that inhibits proliferation of an  
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 XX CC compound's activity; (11) a culture comprising strains in which the gene  
 XX CC product is overexpressed or underexpressed; (12) determining the extent  
 XX CC to which each of the strains is present in a culture or collection of  
 XX CC strains; or (13) identifying the target of a compound that inhibits the  
 XX CC proliferation of an organism. The antisense nucleic acids are useful for  
 XX CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 322 AA;

Query Match 8.8%; Score 92; DB 6; Length 322;  
 Best Local Similarity 22.5%; Pred. No. 2.1;  
 Matches 38; Conservative 34; Mismatches 71; Indels 26; Gaps 7;

QY 32 NSFYATGTAQAVSEPIDVSHLGSITPAAGAGQSDDIGYVAIVKQVNDVKLVK-TLRN 90  
 Db 67 NDNYSTDTSGSYDEPALVKEDRSTISDNK-----LIDKIQLQKEIQELRG 113  
 QY 91 AEOLKPY-FYLIQITSGYETNSTALGNFSETKKAVISLNDNPSAVIL---DKEDIAVLY 146  
 Db 114 QLEVOAHLKLLQQQOVAFYKDLDSRLSNSSTSAKTQNDKPKATDISLGSNSPFTLKVAS 173  
 QY 147 PD-KTGYTNTSIWPGEPDKIIVYNETKPV-ALINFKAFYEAKGMLFD 193  
 Db 174 PQIKAGSSN-----GKQPFVAVSRANPADEQISYLAAYELVKNKRYD 216

## RESULT 13

ABP30785  
 ID ABP30785 standard; protein; 255 AA.

XX AC ABP30785;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polypeptide SEQ ID NO 10746.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus agalactiae.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB004789.

XX PR 27-OCT-2000; 2000GB-00026333.

XX PR 24-NOV-2000; 2000GB-00028727.

XX PR 07-MAR-2001; 2001GB-00005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Massignani V, Margarit Y Rosl, Grandi G, Fraser C;  
 XX PI Tettelin H;

XX DR WPI; 2002-352536/38.

XX DR N-PSDB; ABN71416.

XX PT New Streptococcus protein for the treatment or prevention of infection or  
 XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 XX PT detecting a compound that binds to the protein.

XX PS Claim 1; Page 4181; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 XX CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins

XX SQ Sequence 255 AA;  
 Query Match 8.7%; Score 91; DB 5; Length 255;  
 Best Local Similarity 24.5%; Pred. NO. 1.8;  
 Matches 52; Conservative 31; Mismatches 85; Indels 44; Gaps 10;

QY 5 TLAIAGIIASAAALLAGFATTQSPNSFYATGTAQAVSEPIDVES-HLGSITPAAQAQ 63  
 DB 12 TLVAGALLFSPAVLEVHADQVTTQVNVNNSNQAQMAQKLDQDSIQLENIKD--NVQ 69  
 QY 64 GSD---DIGYAIWIKQVNDVKKVTLR-----NAEQKPKYFKVLQIQIT 106  
 DB 70 GTDYKTVNEAITSVE-----KLKTSLRANPETVYDLSNLSGSRVEALTDVIE--AITFS 121  
 QY 107 SGYETNSTALGN-----PSETKAVISLDNPSAVI-----VLDKEDIAVLVYDKTGYT 153  
 DB 122 TQHLANKVQANIDMGGITKLVIRILDPFASVDSIKRAQVNDVKALEQKLVTPDLKPTD 181  
 QY 154 NTSIWPVGEPPKIIIVYNE--TKPVAILNFKAF 183  
 DB 182 RATIYTKSKUDKEI-WNTRFRDRKKVLNVKEF 212

RESULT 14  
 ADCS1492  
 ID ADCS1492 standard; protein; 978 AA.  
 AC ADCS1492;  
 DT 18-DEC-2003 (first entry)  
 XX Bacterial beta-hexosaminidase gene SEQ ID NO:8.  
 XX beta-hexosaminidase; beta-glucosidase; sphingo-glycolipid.  
 XX Paenibacillus sp.  
 OS JP2003061663-A.  
 XX 04-MAR-2003.  
 XX 21-AUG-2001; 2001JP-00249782.  
 XX 21-AUG-2001; 2001JP-00249782.  
 XX (SANG-) SANGAKU RENKEI KIKO KYUSHU KK.  
 XX WPI; 2003-691955/66.  
 XX N-PSDB; ADCS1491.

PT Novel exo-type ganglioside degradation enzymes e.g. beta- glucosidase and  
 PT beta-hexosaminidase, useful for producing sphingoglycolipids, are derived  
 PT from Paenibacillus species of strain TS12 FERM P-18416.  
 XX Claim 4; SEQ ID NO 8; 23pp; Japanese.

XX The invention relates to a novel recombinant polypeptide, encoded by beta  
 CC - hexosaminidase gene derived from Paenibacillus sp. or encoded by beta-  
 CC glucosidase gene derived from Escherichia coli. A gene of the invention

CC is useful for producing sphingo- glycolipids. The present sequence is  
 CC used in the exemplification of the invention.  
 XX SQ Sequence 978 AA;

Query Match 8.4%; Score 87.5; DB 7; Length 978;  
 Best Local Similarity 25.6%; Pred. No. 28;  
 Matches 43; Conservative 17; Mismatches 69; Indels 39; Gaps 6;  
 QY 48 DVESHLSITPAAAGAGSDDIGYAIWIKQVNDVKKVTLRANAEQLKPYFKYLIQIITS 107  
 DB 477 DFKERLIEHPRLQNGKIKFFADEIIV-----ELPIQVINSEWKMDEGTGTWVVDKTS 528  
 QY 108 GYETNSTALGNFSETKAV-----ISLDNPSAVIVLKDEDIAVLVYDKTGYNTSIVWPGEP 163  
 DB 529 GY-LNGTLVGAKWTAGKQGVSGVDFGSGVINGGQDI-----IGNWTAAVVYVQGP 580  
 QY 164 D-----KIIYVNETKPVAILNFKAFYEAKEGMLPDSLP 196  
 DB 581 NNTNNETLLSGTTSALKINQNTKTKVGI----TIYTKDYTYNYSIP 624

RESULT 15  
 ABB57211  
 ID ABB57211 standard; protein; 284 AA.  
 XX AC ABB57211;  
 XX 07-MAR-2002 (first entry)  
 DT Mouse ischaemic condition related protein sequence SEQ ID NO:514.  
 DE Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 XX vasospastic ischaemia; ischaemic condition; ischaemic disease.  
 XX Mus musculus.  
 XX WO200108188-A2.  
 XX 22-NOV-2001.  
 XX 18-MAY-2001; 2001WO-JP004192.  
 XX 18-MAY-2000; 2000JP-00145977.  
 XX (UYN-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 DR WPI; 2002-034733/04.  
 DR N-PSDB; ABI99523.

Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.

XX Claim 2; Page 1418-1419; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (i) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (i). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention

```
XX
SQ      Sequence 284 RA;
      Query Match      8.3%; Score 86.5; DB 5; Length 284;
      Best Local Similarity 26.2%; Pred. No. 6;
      Matches 45; Conservative 28; Mismatches 68; Indels 31; Gaps 8;

QY      1 VKYTTLAIAGIIASAAALALLAGFATTQSPINSFYATGT--AAQVSEPIDVESHLGSIT 57
Db      126 VQNTATLAVANITNADSATRELLA-----QITLRN--ALGTNLSQIILSDREEIAHHMQSTL 178

QY      58 PAAGAGQSDDIGYAIWMIKDQVNDVKVTLRMAEQLKPFYKYLQITSGYETNSTALG 117
Db      179 DDA---TDDWGIKVERV--EIKDVKLPVQLQRA-----MAAEAEAREARAKVIA 223

QY      118 NFSETKAVISLDNPSAVIIVLDKEDIAVLPDKTGYTNTSIWVPGEPDKLIIVY 169
Db      224 AEGEMNASRALKXASWVIT---ESPAAL---QLRYLQTLTTIAAEKNSTIVF 269
```

Search completed: April 6, 2004, 12:38:53  
Job time : 62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 11:43:46 ; Search time 22 Seconds  
(without alignments)  
485.753 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTTLATAGIIASAAALAL.....EGMLFDSLPIVFNQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	10.1	651	4	US-09-543-681A-8284
2	94	9.0	208	4	US-09-252-991A-32467
3	83	8.0	626	4	US-09-485-717-2
4	80.5	7.7	225	6	5455030-5
5	80.5	7.7	256	4	US-09-107-532A-4072
6	79.5	7.6	769	3	US-09-320-878-12
7	79.5	7.6	769	4	US-09-141-908-10
8	79.5	7.6	769	4	US-09-657-440-12
9	79.5	7.6	809	3	US-09-105-537-24
10	79.5	7.6	1461	4	US-09-976-594-531
11	79.5	7.6	3782	3	US-09-105-537-4
12	79	7.6	1180	3	US-09-224-024-28
13	79	7.6	1180	5	PCT-US94-07902-28
14	78.5	7.5	355	4	US-09-489-039A-12174
15	78	7.5	408	4	US-09-489-039A-13379
16	78	7.5	455	4	US-09-711-164-416
17	77	7.4	555	4	US-09-489-039A-13382
18	77	7.4	927	4	US-09-841-786-3
19	77	7.4	3241	4	US-09-841-786-1
20	76.5	7.3	225	4	US-09-543-681A-4987
21	76.5	7.3	383	3	US-09-413-814-12
22	76.5	7.3	616	1	US-08-385-370-2
23	76.5	7.3	616	1	US-08-385-370-4
24	76.5	7.3	655	1	US-08-469-202-27
25	76.5	7.3	655	2	US-08-484-434C-34
26	76.5	7.3	655	4	US-09-384-361-34
27	76.5	7.3	672	4	US-09-489-039A-12101

28 76 7.3 569 4 US-09-071-035-264 Sequence 264, App  
29 76 7.3 1026 1 US-08-194-290-7 Sequence 7, Appli  
30 76 7.3 1026 3 US-08-614-377A-7 Sequence 7, Appli  
31 76 7.3 1026 3 US-09-142-648B-7 Sequence 7, Appli  
32 76 7.3 1638 4 US-09-071-035-258 Sequence 258, App  
33 76 7.3 1638 4 US-09-071-035-262 Sequence 262, App  
34 76 7.3 1638 4 US-09-071-035-266 Sequence 266, App  
35 76 7.3 1747 4 US-09-134-000C-5999 Sequence 5999, Ap  
36 75.5 7.2 274 4 US-09-813-659-30 Sequence 30, Appl  
37 75.5 7.2 274 4 US-09-549-067A-30 Sequence 30, Appl  
38 75.5 7.2 291 4 US-09-543-681A-6165 Sequence 6165, Ap  
39 75.5 7.2 302 1 US-08-121-054C-30 Sequence 30, Appl  
40 75.5 7.2 302 3 US-08-539-436-30 Sequence 30, Appl  
41 75.5 7.2 302 4 US-09-813-659-32 Sequence 32, Appl  
42 75.5 7.2 302 4 US-09-549-067A-32 Sequence 32, Appl  
43 75.5 7.2 399 4 US-09-252-991A-20610 Sequence 20610, A  
44 75.5 7.2 728 4 US-09-252-991A-28459 Sequence 28459, A  
45 75.5 7.2 1251 4 US-09-252-991A-17263 Sequence 17263, A

## ALIGNMENTS

### RESULT 1

US-09-543-681A-8284  
; Sequence 8284, Application US/09543681A  
; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709,1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; PRIOR FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 8284

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-8284

Query Match 10.1%; Score 105; DB 4; Length 651;  
Best Local Similarity 22.0%; Pred. No. 0.013; 89; Indels 54; Gaps 7;  
Matches 48; Conservative 27; Mismatches 21

QY 14 SAAALALLAGFATQSPPLNSFYATGTAQAVSEPIDVESHLSGIT-PAAGAOGSDDIGYAI 72  
DB 130 SRHVLILIDGVRTNQAGISGVDSQL-----PLSLVQRIEYIRGPRNAVYGSDAIGGVI 184

QY 73 VWIKQDNDVKLVTLRNABQLKPYKYL-QIQTSGYETNSTALGNFSETKAVISLNDP 131  
DB 185 NFITKRPND---GLTNAGIGSHGYQNYNGSIQKVGERITLSAAGGYTKGIHATD-- 239

QY 132 SAVIVLDKEDIAVLPDKTGYNTSIW-----VPE 162  
DB 240 -----DSVPAAFPDHRGFNKSYYWRGIDHIFNDFQGMKGYGYNRTTYVASYPGN 293

QY 163 PDKIIVNETKPAIILNFKAIFYEAKGMLFDSLPIVF 200  
DB 294 TDQSKTYNR-----NVEAGIKFSGDSYFSLITVYS 324

### RESULT 2

US-09-252-991A-32467

; Sequence 32467, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32467  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-252-991A-32467

Query Match 9.0%; Score 94; DB 4; Length 208;  
Best Local Similarity 26.6%; Pred. No. 0.035;  
Matches 46; Conservative 27; Mismatches 62; Indels 38; Gaps 8;

QY 9 AGIASAALALLAGFATTOSPLNSFYATGTAQAVSEPIDV-BSHLGSI-----PA 59  
Db 37 ARTAALGILLPLAAGSAVAEEGSSVKDAKA-AVSSAIEGKLLGGVSEGITSGRQSA 95  
QY 60 AGAGG-----SDDIGYAIWIKQVNDVKLVTLRNAEQLPKPYKYLQIQTSGYETNSTA 115  
Db 96 QGADGAKVSDNAGFA-----EIQQVEVLKTEPRDQQLAVTLGFMKDSDA 140  
QY 116 ---LGNFSETKAVISLNDP-----SAVIVLQKEDIIVLPDKTGYNTSIVWPG 161  
Db 141 TVRLINLTQICALLVIDNGYSNALVALANPDVV--PAKAGIRQTFVPEGG 191

RESULT 3  
US-09-485-717-2  
; Sequence 2, Application US/09485717  
; Patent No. 6673353  
; GENERAL INFORMATION:  
; APPLICANT: Kaufmann, Stefan  
; APPLICANT: Hess, Jirgen  
; TITLE OF INVENTION: Tuberculosis Vaccine  
; FILE REFERENCE: 16862PUS  
; CURRENT APPLICATION NUMBER: US/09/485,717  
; CURRENT FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: EP 97114614.7  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: PCT/EP98-05109  
; PRIOR FILING DATE: 1998-08-12  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 626  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant  
; US-09-485-717-2

Query Match 8.0%; Score 83; DB 4; Length 626;  
Best Local Similarity 19.5%; Pred. No. 3.2;  
Matches 43; Conservative 33; Mismatches 71; Indels 74; Gaps 7;

QY 11 IIASAALAL-----LAGFATT-----QSPLNSFYATGTAQAVSEPI 47  
Db 16 MICTAAAVVLPGLVGLAGGAATAGAFSRPGLPVEYLOSARQSAANKLHSAQSTKQASAF 75  
QY 48 DVESHLSGSIPTAAGAGQSDDIGVAIVWIKQVNDVKLVTLRNAEQLPKPYKYL----- 101  
Db 76 NKENSISMAPPPASPPASP-----KTPIEKKHADEIKYIQGLDYNKNN 119  
QY 102 -----QIQTSGYETNSTAL-----GNFSETKAVISLNDPNSAVIVLD 138  
Db 120 VLVVHGDAVNTVPKPKYKGDNEIIVVEKKKSSINONNADIQVNAISSITYPGALVKAN 179  
QY 139 KEI-----AVLYPKTGYNTSIVWPG-----BPDKIIVNETK 173

Db 180 SELVENQPDVLPVKRSLTILSLDLPMTNQDNKIVVKNATK 220

RESULT 4  
5455030-5  
; Patent No. 5455030  
; APPLICANT: LADNER, ROBERT C.; BIRD, ROBERT E.; HARDMAN, KARL  
; TITLE OF INVENTION: IMMUNOTHERAPY USING SINGLE CHAIN  
; POLYPEPTIDE BINDING MOLECULES  
; NUMBER OF SEQUENCES: 24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/40,440  
; FILING DATE: 1-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 512,910  
; FILING DATE: 25-APR-1990  
; APPLICATION NUMBER: 299,617  
; FILING DATE: 19-JAN-1989  
; APPLICATION NUMBER: 92,110  
; FILING DATE: 02-SEP-1987  
; APPLICATION NUMBER: 902,971  
; FILING DATE: 01-SEP-1986  
; SEQ ID NO: 5  
; LENGTH: 225  
5455030-5

Query Match 7.7%; Score 80.5; DB 6; Length 225;  
Best Local Similarity 24.5%; Pred. No. 1.2;  
Matches 40; Conservative 22; Mismatches 66; Indels 35; Gaps 8;

QY 22 AGFATTQSPLNSFYATGTAQAVSEPIDVESHLSGSIPTAAGAGQSDDIGVAIVWIKQVND 81  
Db 26 ASSSVSSVYLHWFOQKSIAXAFKNGGDLVPGSLKSCSAAGFTFISGMSWVR-QTPD 84  
QY 82 VKLK--VTLRNAEQLPKPYKYLQIQTSGYETNSTALGNFSETKAVISLNDPNSAVIVLD- 138  
Db 85 KRLEWVATISSGS-----TYTYYPDSVK-GRFT-----ISRDNAKDTLYLQM 125  
QY 139 ----KEDIAVLPDK-----TGYNTSIVWPGEPDKIIVN 170  
Db 126 SGLKSEDTAMYYCARRITTVLTLTDYYANDYWGPGSP-KLWIYS 167

RESULT 5  
US-09-107-532A-4072  
; Sequence 4072, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FABRIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:

/ NAME: Ariniello, Pamela Deneke  
/ REGISTRATION NUMBER: 40,489  
/ REFERENCE/DOCKET NUMBER: GTC-012  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (781)893-5007  
/ TELEFAX: (781)893-8277  
/ INFORMATION FOR SEQ ID NO: 4072:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 256 amino acids  
/ TYPE: amino acid  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ HYPOTHETICAL: YES  
/ ORIGINAL SOURCE:  
/ ORGANISM: Enterococcus faecium  
/ FEATURE:  
/ NAME/KEY: misc-feature  
/ LOCATION: (B) LOCATION 1...256  
/ SEQUENCE DESCRIPTION: SEQ ID NO: 4072:  
/ US-09-107-532A-4072

Query Match 7.7%; Score 80.5; DB 4; Length 256;  
Best Local Similarity 25.7%; Pred. No. 1.5;  
Matches 43; Conservative 24; Mismatches 61; Indels 39; Gaps 8;  
QY 44 SEPIDVESHLSITPAAGAGSDDIGYAIWIKD--QVNDVKLVIRNAEQKPYFKYL 101  
DB 91 SEQLDNQGH-----EKQGVNGAKFSVVDSDIILQKMDVKDLTTDQIESQLKDRVKKL 143  
QY 102 ---QIQTSGYETNS-----TALGNFS-----ETKAVISLDNPSNAVILDKEDIAVLYP- 147  
DB 144 SSDQLKLVNGETKTDIQQTGTFEFSVEQANGKQAYIIVNESPENISNSEDILLTPV 203  
QY 148 -DKTGYNTSIWV-----PGEPKIIVYNETKPVAILNFKAFYE 185  
DB 204 SDKNGELKDVWLYPKSEASQPKVEVKIVSTGVK-----KNFFE 243

RESULT 6  
US-09-320-878-12  
/ Sequence 12, Application US/09320878A  
/ Patent No. 6117659  
/ GENERAL INFORMATION:  
/ APPLICANT: ASHLEY, Gary  
/ APPLICANT: BETLACH, Melanie C.  
/ APPLICANT: BETLACH, Mary C.  
/ APPLICANT: MCDANIEL, Robert  
/ APPLICANT: TANG, Li  
/ TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
/ FILE REFERENCE: 300622002120  
/ CURRENT APPLICATION NUMBER: US/09/320,878A  
/ CURRENT FILING DATE: 1999-05-27  
/ EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
/ EARLIER FILING DATE: 1998-08-28  
/ EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
/ EARLIER FILING DATE: 1998-05-06  
/ EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
/ EARLIER FILING DATE: 1997-04-30  
/ EARLIER APPLICATION NUMBER: 60/119,139  
/ EARLIER FILING DATE: 1999-02-08  
/ EARLIER APPLICATION NUMBER: 60/100,880  
/ EARLIER FILING DATE: 1998-09-22  
/ EARLIER APPLICATION NUMBER: 60/087,080  
/ EARLIER FILING DATE: 1998-05-28  
/ NUMBER OF SEQ ID NOS: 34  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 12  
/ LENGTH: 769  
/ TYPE: PRT  
/ ORGANISM: Streptomyces venezuelae  
/ US-09-320-878-12

Query Match 7.6%; Score 79.5; DB 3; Length 769;

Best Local Similarity 19.8%; Pred. No. 11;  
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;  
QY 17 ALALLAGFATTQSPPLNSFYATGTAQAVSEP-IDVESHLSITPAAGAGSDDIGYAIWVI 75  
DB 428 AVRATGGYATVOLGSHTEIAGQVYGVKSSPLLKLTGTHKLTISGFAMSGATPLSLELGV 487  
QY 76 KQVNDVKLVIRNAEQKPYFKYLQIQTSGYETNSTAL-GNFSETKAVISLDNPSAV 134  
DB 488 TPAADATTAKAVESARKARTAVFAYDDGTGTEGVRPNLSLPGTQDKLISAVADANPNTI 547  
QY 135 IVLDKEDIAVLYP--DKTGYNTSIWVPG-----EPDKIIVYNETKPVAILNFKAFYEAK 188  
DB 548 VVLTGSG-SVLMFWLSKTRAV-LDMWYPCQAGAEATAALLYGDVNP-----S 592  
QY 189 GMLFDSLPIVFN 200  
DB 593 GKLTQSFPAAEN 604

RESULT 7  
US-09-141-908-10  
/ Sequence 10, Application US/09141908  
/ Patent No. 6503741  
/ GENERAL INFORMATION:  
/ APPLICANT: ASHLEY, Gary  
/ APPLICANT: BETLACH, Melanie C.  
/ APPLICANT: BETLACH, Mary  
/ APPLICANT: MCDANIEL, Robert  
/ APPLICANT: TANG, Li  
/ TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
/ FILE REFERENCE: 300622002100  
/ CURRENT APPLICATION NUMBER: US/09/141,908  
/ CURRENT FILING DATE: 1998-08-28  
/ EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
/ EARLIER FILING DATE: 1998-05-06  
/ EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
/ EARLIER FILING DATE: 1997-04-30  
/ EARLIER APPLICATION NUMBER: PROV. 60/076,919  
/ EARLIER FILING DATE: 1998-03-05  
/ EARLIER APPLICATION NUMBER: PROV. 60/087,080  
/ EARLIER FILING DATE: 1998-05-28  
/ NUMBER OF SEQ ID NOS: 31  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 10  
/ LENGTH: 769  
/ TYPE: PRT  
/ ORGANISM: Streptomyces venezuelae  
/ US-09-141-908-10

Query Match 7.6%; Score 79.5; DB 4; Length 769;  
Best Local Similarity 19.8%; Pred. No. 11;  
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;  
QY 17 ALALLAGFATTQSPPLNSFYATGTAQAVSEP-IDVESHLSITPAAGAGSDDIGYAIWVI 75  
DB 428 AVRATGGYATVOLGSHTEIAGQVYGVKSSPLLKLTGTHKLTISGFAMSGATPLSLELGV 487  
QY 76 KQVNDVKLVIRNAEQKPYFKYLQIQTSGYETNSTAL-GNFSETKAVISLDNPSAV 134  
DB 488 TPAADATTAKAVESARKARTAVFAYDDGTGTEGVRPNLSLPGTQDKLISAVADANPNTI 547  
QY 135 IVLDKEDIAVLYP--DKTGYNTSIWVPG-----EPDKIIVYNETKPVAILNFKAFYEAK 188  
DB 548 VVLTGSG-SVLMFWLSKTRAV-LDMWYPCQAGAEATAALLYGDVNP-----S 592  
QY 189 GMLFDSLPIVFN 200  
DB 593 GKLTQSFPAAEN 604

RESULT 8





NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 3782  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-4

Query Match 7.6%; Score 79.5; DB 3; Length 3782;  
Best Local Similarity 19.8%; Pred. No. 1.3e+02;  
Matches 38; Conservative 34; Mismatches 97; Indels 23; Gaps 7;  
Db

QY 17 ALALAGAFATQSPNSFYATGTAQAVSEP-IDVESHLSGITPAAGAGQSGDDIGYAIWVI 75  
Db 1552 AVRATGGYATVQLGSHTEAGVYGVKSSPLLKLTGTHKLTISGFAMSNATPLSLELGV 1611  
QY 76 KDQVNDVKLVTURNABQLKPYFYKIQITSGYETNSTAL-GNFSETKAVISLDNPSAV 134  
Db 1612 TPAADATTIAKAVESARKARTAVVFAYDGTGVDPRNLSLPGTKLISAVADANPNTI 1671  
QY 135 IVLDKEDIAVLXP-DKTGYTNTSIWVPG---EPDKIIIVNETKPVAILNFKAFYEAKE 188  
Db 1672 VLVNTGS-SVLMPLWLSKTRAV-LDMWYPGQAGAEATAALLYGDVNP-----S 1716  
QY 189 GMLFDSLPIVFN 200  
Db 1717 GKLTQSFRAEN 1728

RESULT 12  
US-09-224-024-28  
; Sequence 28, Application US/09224024  
; Patent No. 6056953  
; GENERAL INFORMATION:  
; APPLICANT: Leslie Hickie  
; APPLICANT: Jewel Payne  
; TITLE OF INVENTION: Materials and Methods for the Control of  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/224,024  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/856,226  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA79  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1180 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-224-024-28

Query Match 7.6%; Score 79; DB 3; Length 1180;  
Best Local Similarity 22.2%; Pred. No. 23;  
Matches 47; Conservative 31; Mismatches 66; Indels 68; Gaps 10;  
Db

QY 24 FATTQS---PLNSFYATGTAQAVSEPIDVESHLSGITPAAGAGQSGDDIGYAIWIKDQVN 80  
Db 520 FAWTHSSVDPKNTIYTHLTQI---PAVKANSLG---TASKVVQGPCHTGGDLIDFKDH-- 572  
QY 81 DVKLKVTLENAQLKPYFYKIQITSGYETNSTALGNFSETKAVISLDNP----- 131  
Db 573 ---FKITQHSNFQOSYF-----IRIRYASNGSA-----NTRAVINLSIPGVAEIGMAL 618  
QY 132 -----SAVIVLDKEDIAVLXPDKTGYTNTSIWVPGEPDKIIV 168  
Db 619 NTFSGTDYTNLYKXDFQYLEFSNEVKFAPNQNISLVFNESDVYNTTTLI-----DKI-- 672  
QY 169 YNETKPVAILNFKAFYEAKGMLFDSLPIVFN 200  
Db 673 --EFLPIT---RSIREDEKQKLETVOQIIN 698

RESULT 13  
PCT-US94-07902-28  
; Sequence 28, Application PC/TUS9407902  
; GENERAL INFORMATION:  
; APPLICANT: Street address: 4980 Carroll Canyon Road  
; APPLICANT: City: San Diego  
; APPLICANT: State/Province: California  
; APPLICANT: Country: US  
; APPLICANT: Postal code/Zip: 92121  
; APPLICANT: Phone number: (619) 453-8030  
; APPLICANT: Telex number:  
; TITLE OF INVENTION: Materials and Methods for the Control of  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07902  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: MA79  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1180 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US94-07902-28

Query Match 7.6%; Score 79; DB 5; Length 1180;  
Best Local Similarity 22.2%; Pred. No. 23;  
Matches 47; Conservative 31; Mismatches 66; Indels 68; Gaps 10;  
Db

```
QY 24 FATTQS---PLNGFYATGTAQVSEPIDVESHLSITPAAGAGSGDDIGYAIWIKDOVN 80
Db 520 FATHSSVDKNYIYTHLTQI---PAVKANSIG--TASKVQPGHGTGDLIDFKDH-- 572
QY 81 DVKLKVTLRNAQLKPYKYLQIQITSGYETNSTALGNFSETKAVISLNDP----- 131
Db 573 ---FKITCOHSNFQOSYF-----IRRYASNGSA-----NTRAVINLSIPGVAELGMAL 618
QY 132 -----SAVIVLDKEDIYLYPKTGYTNTSIWVPGEPDKIIV 168
Db 619 NPFTSGTDYNNLYKDFQYLEFSNEVKFAPQNLSLVNRSVDYTNVTLLI---DKI-- 672
QY 169 YNETKPVAILNFKAFYEAKEGMLPDSLPIVFN 200
Db 673 --BFLPIT---RSIREDKQKLETVQQIIN 698

RESULT 14
US-09-489-039A-12174
; Sequence 12174, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12174
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12174

Query Match 7.5%; Score 78.5; DB 4; Length 355;
Best Local Similarity 25.2%; Pred. No. 4;
Matches 57; Conservative 33; Mismatches 85; Indels 51; Gaps 13;

QY 6 LTAIAGIITASAAALALLAGFATTQSPPLNSFYA---TGTAQVSEPIDVE-SHLGSIITPAAG 61
Db 36 LTVAAALSGSALAEVRIALVAKSLGNGFFEAANVGAQAQAKELGDVKVIYTGPTTTTAAE 95
QY 62 -----AQSGDDIGYAIWIKDOVNDVKLV-TLRNAEQ-----LKPYFK 99
Db 96 AQIDVNLGLIAQGVDAIAIS-----ANDPDVFPVKRAMQGIKIVSWDSGVAPAGR 148
QY 100 YLQIQTSGYETNSTALGNFSETKAVISLNDPNSAVIVLDKEDIAVLYPDKTGYTNTSIWV 159
Db 149 QIHLEN-----PSNNALIG---ETNVKLAADALQALNV-EKGEVAVLSATPTS-TNQNIWI 198
QY 160 PGEFDKLIIVNETKPVAI-----LNFKAFYEAKEGML---FDSLPIVI 198
Db 199 EEMKKVLPQVPSVQLVTVAYGDDLSKSYREA-VGLLKSYPDLKVI 243

RESULT 15
US-09-489-039A-13379
; Sequence 13379, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13379
```

```
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13379

Query Match 7.5%; Score 78; DB 4; Length 408;
Best Local Similarity 22.2%; Pred. No. 5.7;
Matches 69; Conservative 31; Mismatches 73; Indels 138; Gaps 17;

QY 1 VKT-----TIAIAGIITASAAALALLAGFATTQSPPLNSFYATGTAQVSEPIDVESHU--- 53
Db 49 MKYTLPALTAIS-----AALSGCATPHS-----SAVSQPV-VDSPVPNV 87
QY 54 -----GSITPAAGAGSGDDIGYAIWIKD--QVNDVKLV-TLRN 90
Db 88 AQPLQQLAEGLYEMALSFGDDALYVASAEGFKNVQGGAVYLLDPHTLNTIGLTHTDLKN 147
QY 91 -AEQLKPYFK--YLQIQTSGYETNSTALGN-----FSE-----TKAVISLND-- 130
Db 148 FALQLSAEGKTLVSNLSLDGGISAITATGKVKNRLLFSENEKGRPYGARQALLNNTL 207
QY 131 -----PSAVIVLDKEDIAVLYPDKTGYTNTSIWVPG-----EPDKIIVY 169
Db 208 YVGAVADPAQIWIYVDATTLKL-----KTRIKNTGKWTGLHYSATQGRVVAANGSEILVI 263
QY 170 NE-----TKPVAILNFKAFYEAKEGMLF----- 192
Db 264 NPNRQRIEQRWKPLGDKPALLNMAE--DSDTGRLFVTDNSKAKTTLVLDIHSGKLLKQL 321
QY 193 ---DSLPIVFN 200
Db 322 DVGDSLAVQFN 332

Search completed: April 6, 2004, 12:33:56
Job time : 23 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:33:31 ; Search time 42. Seconds  
(without alignments)  
1294.407 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTLAIGIATASAAALAL.....EGMLFDSLPIVFNFQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1042	100.0	207	10	US-09-997-807-2
2	1042	100.0	207	15	US-10-370-370-2
3	553.5	53.1	170	10	US-09-997-807-4
4	553.5	53.1	170	15	US-10-370-370-4
5	540	51.8	178	10	US-09-997-807-6
6	540	51.8	178	15	US-10-370-370-6
7	473	45.4	140	15	US-10-370-370-12
8	307.5	29.5	131	15	US-10-370-370-8
9	305.5	29.3	124	10	US-09-997-807-10
10	305.5	29.3	124	15	US-10-370-370-10
11	304.5	29.2	130	10	US-09-997-807-8
12	96	9.2	235	14	US-10-134-021-4
13	92	8.8	322	12	US-10-282-122A-61228
14	86.5	8.3	846	10	US-09-893-519A-34
15	86.5	8.3	846	15	US-10-369-493-1972

16 85.5 8.2 337 12 US-10-412-699B-1672 Sequence 1672, Ap  
17 85.5 8.2 337 15 US-10-374-780A-1646 Sequence 1646, Ap  
18 85.5 8.2 337 15 US-10-072-012-415 Sequence 415, App  
19 84.5 8.1 506 15 US-10-108-260A-4387 Sequence 4387, App  
20 84.5 8.1 1886 12 US-10-147-239A-6 Sequence 6, Appli  
21 83 8.0 626 9 US-09-948-722-2 Sequence 2, Appli  
22 83 8.0 1246 9 US-09-919-497-85 Sequence 85, Appl  
23 83 8.0 1247 10 US-09-961-403-14 Sequence 14, Appl  
24 82.5 7.9 243 14 US-10-171-404A-42 Sequence 42, Appl  
25 82.5 7.9 243 15 US-10-374-780A-2004 Sequence 2004, Ap  
26 82.5 7.9 296 12 US-10-282-122A-52171 Sequence 52171, A  
27 82.5 7.9 1954 12 US-10-147-239A-4 Sequence 4, Appli  
28 81.5 7.8 320 12 US-10-335-977-5813 Sequence 5813, Ap  
29 81.5 7.8 349 12 US-10-335-977-5814 Sequence 5814, Ap  
30 81 7.8 530 12 US-10-424-599-262050 Sequence 262050,  
31 81 7.8 751 14 US-10-032-585-7803 Sequence 7803, Ap  
32 80.5 7.7 258 9 US-09-738-626-6436 Sequence 6436, Ap  
33 80.5 7.7 292 12 US-10-424-599-240558 Sequence 240558,  
34 80 7.7 398 12 US-10-282-122A-48705 Sequence 48705, A  
35 80 7.7 612 12 US-10-282-122A-64164 Sequence 64164, A  
36 80 7.7 1647 14 US-10-032-585-7572 Sequence 7572, Ap  
37 79.5 7.6 317 15 US-10-369-493-6984 Sequence 6984, Ap  
38 79.5 7.6 447 12 US-10-282-122A-51512 Sequence 51512, A  
39 79.5 7.6 769 10 US-09-793-708-12 Sequence 12, Appl  
40 79.5 7.6 769 14 US-10-201-365-10 Sequence 10, Appl  
41 79.5 7.6 769 14 US-10-160-539-12 Sequence 12, Appl  
42 79.5 7.6 809 9 US-09-861-289-24 Sequence 24, Appl  
43 79.5 7.6 809 9 US-09-860-846-24 Sequence 24, Appl  
44 79.5 7.6 809 10 US-09-988-384B-24 Sequence 24, Appl  
45 79.5 7.6 809 10 US-09-836-821-24 Sequence 24, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-997-807-2  
; Sequence 2, Application US/09997807  
; Publication No. US20030198641  
; GENERAL INFORMATION:  
; APPLICANT: Jay Short  
; APPLICANT: Eric J. Mathur  
; APPLICANT: W. Michael Laferty  
; APPLICANT: Nelson Barton  
; APPLICANT: Kevin Chow  
; TITLE OF INVENTION: Method of Making A Protein Polymer and  
; TITLE OF INVENTION: Uses of the Polymer  
; FILE REFERENCE: DVSA-1005US  
; CURRENT APPLICATION NUMBER: US/09/997,807  
; PRIOR FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: 60/250,426  
; PRIOR FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 207  
; TYPE: PPT  
; ORGANISM: Pyrodictium abyssi  
US-09-997-807-2

Query Match 100.0%; Score 1042; DB 10; Length 207;  
Best Local Similarity 100.0%; Pred. No. 1.8e-101;  
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTLAIGIATASAAALALAGFATTQSPFLNSFVATGTAAQVSEPIDVESHLSITPAA 60  
Db 1 VKYTLAIGIATASAAALALAGFATTQSPFLNSFVATGTAAQVSEPIDVESHLSITPAA 60  
QY 61 GAQSDDIGYAIWIKQVNDVKLVKTLRNALQKPKFYKLOITSGYETNSTALGNPS 120  
Db 61 GAQSDDIGYAIWIKQVNDVKLVKTLRNALQKPKFYKLOITSGYETNSTALGNPS 120  
QY 121 ETKAVISLONPSAVIILVDKEDIADVLPDKTGYTNTSIWVPGEPDKIIVYNKTPVAILNF 180

```
|||||
Db 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207
Db 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207

RESULT 2
US-10-370-370-2
; Sequence 2, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: THEM AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-10-370-370-2

Query Match 100.0%; Score 1042; DB 15; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.8e-101;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKYTTLAIAGIIASAALALLAGFATQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
Db 1 VKYTTLAIAGIIASAALALLAGFATQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
QY 61 GAQGSDDIGYAIWIWKQDVNDVKLVTLRNAEOLKPYFKYLQITSGYETNSTALGNFS 120
Db 61 GAQGSDDIGYAIWIWKQDVNDVKLVTLRNAEOLKPYFKYLQITSGYETNSTALGNFS 120
QY 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
Db 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207
Db 181 KAFYEAKGMLFDSLPIVFNQVLQVG 207

RESULT 3
US-09-997-807-4
; Sequence 4, Application US/09997807
; Publication No. US20030198681A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE OF INVENTION: Uses of the Polymer
; FILE REFERENCE: DVSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-4

Query Match 53.1%; Score 553.5; DB 15; Length 170;
Best Local Similarity 60.2%; Pred. No. 4.4e-50;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;
```

```
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-4

Query Match 53.1%; Score 553.5; DB 10; Length 170;
Best Local Similarity 60.2%; Pred. No. 4.4e-50;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;

QY 1 VKYTTLAIAGIIASAALALLAGFATQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
Db 1 VKPTALALAGIIASAADLALLAGFATQSPPLNSFYATGTAATSEPIDVESHLSITPAA 60
QY 61 GAQGSDDIGYAIWIWKQDVNDVKLVTLRNAEOLKPYFKYLQITSGYETNSTALGNFS 120
Db 61 GAQGSDDIGYFNVTAKDQVNVTKIYTLANAEOLKPYFKYLQIVLKSEVA-----D 111
QY 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
Db 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 206
Db 144 TAYEAKGMLFDSLPIVFNQVLQV 169

RESULT 4
US-10-370-370-4
; Sequence 4, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Frey, Gerhard
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: THEM AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-10-370-370-4

Query Match 53.1%; Score 553.5; DB 15; Length 170;
Best Local Similarity 60.2%; Pred. No. 4.4e-50;
Matches 124; Conservative 12; Mismatches 33; Indels 37; Gaps 2;

QY 1 VKYTTLAIAGIIASAALALLAGFATQSPPLNSFYATGTAQAVSEPIDVESHLSITPAA 60
Db 1 VKPTALALAGIIASAADLALLAGFATQSPPLNSFYATGTAATSEPIDVESHLSITPAA 60
QY 61 GAQGSDDIGYAIWIWKQDVNDVKLVTLRNAEOLKPYFKYLQITSGYETNSTALGNFS 120
Db 61 GAQGSDDIGYFNVTAKDQVNVTKIYTLANAEOLKPYFKYLQIVLKSEVA-----D 111
QY 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
Db 121 ETKAVISLDNPSAVIVLDKEDIAVLDPKGTNTSIVWVGEPDKIIVNETKPVALLNF 180
QY 181 KAFYEAKGMLFDSLPIVFNQVLQVG 206
Db 144 TAYEAKGMLFDSLPIVFNQVLQV 169

RESULT 5
US-09-997-807-6
; Sequence 6, Application US/09997807
; Publication No. US20030198681A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE REFERENCE: DUSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
; US-09-997-807-6

Query Match
Best Local Similarity 51.8%; Score 540; DB 10; Length 178;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY 1 VKYTLAAGIIVIAAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLS-GSITPA 59
Db 1 MRYTTLAAGIVASAAALALLAGFATTQSPSSFYATGTAQAVSEPIDVESHLDNTIAPA 60
QY 60 AGAGSDDIGYAIWIKQVNDVKLVTLNABQLKPYFKYLIQITSGYETNSTALGNF 119
Db 61 AGAQGYKDMGYIKITQSKVNVIKLVTLNABQLKPYFDYLVLS-----NATGT--- 113
QY 120 SETKAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILN 179
Db 114 DMVKAVLSLEKPSAVIILDND-----YDSTNKIQ-LK 145
QY 180 FKAFYEAKGMLFDSLPIVFNFOVL 204
Db 146 VEAYEAKGMLFDSLPIVFNFOVL 170

RESULT 6
US-10-370-370-6
; Sequence 6, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
; US-10-370-370-6

Query Match
Best Local Similarity 51.8%; Score 540; DB 15; Length 178;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY 1 VKYTLAAGIIVIAAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLS-GSITPA 59
Db 1 MRYTTLAAGIVASAAALALLAGFATTQSPSSFYATGTAQAVSEPIDVESHLDNTIAPA 60
QY 60 AGAGSDDIGYAIWIKQVNDVKLVTLNABQLKPYFKYLIQITSGYETNSTALGNF 119
Db 61 AGAQGYKDMGYIKITQSKVNVIKLVTLNABQLKPYFDYLVLS-----NATGT--- 113
QY 120 SETKAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILN 179
Db 114 DMVKAVLSLEKPSAVIILDND-----YDSTNKIQ-LK 145
QY 180 FKAFYEAKGMLFDSLPIVFNFOVL 204
Db 146 VEAYEAKGMLFDSLPIVFNFOVL 170

RESULT 6
US-10-370-370-6
; Sequence 6, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
; US-10-370-370-6

Query Match
Best Local Similarity 51.8%; Score 540; DB 15; Length 178;
Matches 120; Conservative 22; Mismatches 27; Indels 36; Gaps 5;

QY 1 VKYTLAAGIIVIAAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLS-GSITPA 59
Db 1 MRYTTLAAGIVASAAALALLAGFATTQSPSSFYATGTAQAVSEPIDVESHLDNTIAPA 60
QY 60 AGAGSDDIGYAIWIKQVNDVKLVTLNABQLKPYFKYLIQITSGYETNSTALGNF 119
Db 61 AGAQGYKDMGYIKITQSKVNVIKLVTLNABQLKPYFDYLVLS-----NATGT--- 113
QY 120 SETKAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILN 179
Db 114 DMVKAVLSLEKPSAVIILDND-----YDSTNKIQ-LK 145
QY 180 FKAFYEAKGMLFDSLPIVFNFOVL 204
Db 146 VEAYEAKGMLFDSLPIVFNFOVL 170

RESULT 6
US-10-370-370-6
; Sequence 6, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
; US-10-370-370-6

```

```

QY 60 AGAGSDDIGYAIWIKQVNDVKLVTLNABQLKPYFKYLIQITSGYETNSTALGNF 119
Db 61 AGAQGYKDMGYIKITQSKVNVIKLVTLNABQLKPYFDYLVLS-----NATGT--- 113
QY 120 SETKAVISLDNPSAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILN 179
Db 114 DMVKAVLSLEKPSAVIILDND-----YDSTNKIQ-LK 145
QY 180 FKAFYEAKGMLFDSLPIVFNFOVL 204
Db 146 VEAYEAKGMLFDSLPIVFNFOVL 170

RESULT 7
US-10-370-370-12
; Sequence 12, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Frey, Gerhard
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
; US-10-370-370-12

Query Match
Best Local Similarity 45.4%; Score 473; DB 15; Length 140;
Matches 119; Conservative 7; Mismatches 8; Indels 60; Gaps 8;

```

```

QY 5 TLALAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSITPAAGAC 64
Db 3 TLALAGIIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLSITPAAGAC 59
QY 65 SDDIGYAIWIKQVNDVKLVTLNABQLKPYFKYLIQITSGYETNSTALGNFSETKA 124
Db 60 S-DIGYII-----KVNVLKLVTLNABQLKPYFKYLIQIVLS-----SEIKA 100
QY 125 VISLNPASAVIVLDKEDIADVLPDKTGYTNTSIWVGPDPKIIIVNETKPVAILNFKAFY 184
Db 101 VISLDPKPSAVIILD-EDFAI-----AYY 122
QY 185 EAKGMLFDSLPIV 198
Db 123 EAKGMLFDSLPIV 136

RESULT 8
US-10-370-370-8
; Sequence 8, Application US/10370370
; Publication No. US20040006778A1
; GENERAL INFORMATION:
; APPLICANT: Barton, Nelson R.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Frey, Gerhard
; TITLE OF INVENTION: CHIMERIC CANNULAE PROTEINS, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-115001
; CURRENT APPLICATION NUMBER: US/10/370,370
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 60/357,406

```

```
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-10-370-370-8

Query Match          29.5%; Score 307.5; DB 15; Length 131;
Best Local Similarity 48.1%; Pred. No. 2.6e-24;
Matches 76; Conservative 16; Mismatches 39; Indels 27; Gaps 4;

Qy 33 SFYATGTAQAVSEPIDVESHLSITPAAGAGSDDIGYAIWIKQVNDVKLVTLRNAE 92
Db 1 SFYATGTAQAVSEPIDVSSSLGTLNTAAGAGQKQTLGDIYIAHNDVNIKLKVTLANAA 60

Qy 93 QLKPYFKYLIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIIVLPDKTG 152
Db 61 QLRPYFKYLIKLVS-LDSN-----GNESEKGMITLWKPVAVIILDHEDF----- 105

Qy 153 TNSIWPGEEDKIIVNETKPVAILNFKAFYEAKGML 190
Db 106 -NNDIDNGND-----AKIRVAYYEAKGML 131

RESULT 9
US-09-997-807-10
; Sequence 10, Application US/09997807
; Publication No. US20030198681A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE REFERENCE: DVSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-10

Query Match          29.3%; Score 305.5; DB 10; Length 124;
Best Local Similarity 48.8%; Pred. No. 3.9e-24;
Matches 78; Conservative 14; Mismatches 31; Indels 37; Gaps 5;

Qy 33 SFYATGTAQAVSEPIDVESHLSITPAAGAGSDDIGYAIWIKQVNDVKLVTLRNA 91
Db 1 SFYATGTAATSEPIDVWSNLNTAIPAGAGQSGVIGSITIENKTDVNVVVKLITLANA 60

Qy 92 EQLKPYFKYLIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIIVLPDKTG 151
Db 61 EQLKPYFDYLIQVLKS-VDSN-----EIKAVLSLEKPSAVIILDNEDF-----QG 104

Qy 152 YNTSIWVPGEDKIIVNETKPVAILNFKAFYEAKGML 191
Db 105 GDNQ-----CQIDATAYYEAKGML 124

RESULT 11
US-09-997-807-8
; Sequence 8, Application US/09997807
; Publication No. US20030198681A1
; GENERAL INFORMATION:
; APPLICANT: Jay Short
; APPLICANT: Eric J. Mathur
; APPLICANT: W. Michael Lafferty
; APPLICANT: Nelson Barton
; APPLICANT: Kevin Chow
; TITLE OF INVENTION: Method of Making A Protein Polymer and
; FILE REFERENCE: DVSA-1005US
; CURRENT APPLICATION NUMBER: US/09/997,807
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,426
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Pyrodicticum abyssi
US-09-997-807-8

Query Match          29.2%; Score 304.5; DB 10; Length 130;
Best Local Similarity 47.2%; Pred. No. 5.4e-24;
Matches 75; Conservative 22; Mismatches 33; Indels 29; Gaps 6;

Qy 33 SFYATGTAQAVSEPIDVESHLSITPAAGAGSDDIGYAIWIKQVNDVKLVTLRNAE 92
Db 1 SFYATGTAQAVSEPIDVWSNLNTAAGAGQKQTLGDIYIAHNDVNIKLKVTLANAA 60

Qy 93 QLKPYFKYLIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIIVLPDKTG 152
Db 61 QLRPYFKYLIKLVS-LDSN-----GNESEKGMITLWKPVAVIILDHEDF-----F 105

Qy 153 TNSIWPGEEDKIIVNETKPVAILNFKAFYEAKGML 191
```





```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 107
US-09-893-519A-34

Query Match      8.3%; Score 86.5; DB 10; Length 846;
Best Local Similarity 24.1%; Pred. No. 8.5;
Matches 52; Conservative 38; Mismatches 81; Indels 45; Gaps 12;

QY 25 ATTQSPLNSFYATGTAQAVSEPI---DVESHLSGITPAAGAGSDD-IGYAIWIKQVYN 80
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 ATQLGPLIDFYKQLSDSPETTMMHKIEKFDAL---FGIQTDDMVKYLVSFQSLPS 98
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 DVKLKVTLR---NAEQLKPKYKYLQ-IQITSGY-ETNSTALGNFSETKAVIS----- 127
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 NYRAKIVQKSSGLNMLNLANHEHLLSPVRAPSIYTEASFENMDRPFSESRSMVSSPNRYVP 158
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 -----LDNPSAVIVLDKEDIAVLPDKTGYNTSIWVGEPEDKIIIVYNETKPA 176
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 SSTYSSVTLRQLSNPYVYNTIPEEDI-LKYVSYTLATTSALFPFDHEQIQI-----PSK 212
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 ILNFKA-----FYEAKGMLFDSLPV-IFNFQVLQV 206
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 IPNFESGLHLHIFEA--GLLYQSLGYKVEKFRMLNI 246
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-10-369-493-1972
; Sequence 1972, Application US/10369493
; Publication No. US20030333675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1972
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1972

Query Match      8.3%; Score 86.5; DB 15; Length 846;
Best Local Similarity 24.1%; Pred. No. 8.5;
Matches 52; Conservative 38; Mismatches 81; Indels 45; Gaps 12;

QY 25 ATTQSPLNSFYATGTAQAVSEPI---DVESHLSGITPAAGAGSDD-IGYAIWIKQVYN 80
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 ATQLGPLIDFYKQLSDSPETTMMHKIEKFDAL---FGIQTDDMVKYLVSFQSLPS 98
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 DVKLKVTLR---NAEQLKPKYKYLQ-IQITSGY-ETNSTALGNFSETKAVIS----- 127
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 NYRAKIVQKSSGLNMLNLANHEHLLSPVRAPSIYTEASFENMDRPFSESRSMVSSPNRYVP 158
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 -----LDNPSAVIVLDKEDIAVLPDKTGYNTSIWVGEPEDKIIIVYNETKPA 176
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 SSTYSSVTLRQLSNPYVYNTIPEEDI-LKYVSYTLATTSALFPFDHEQIQI-----PSK 212
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 ILNFKA-----FYEAKGMLFDSLPV-IFNFQVLQV 206
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 IPNFESGLHLHIFEA--GLLYQSLGYKVEKFRMLNI 246
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:37:42 ; Search time 20 Seconds  
(without alignments)  
995.582 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTTLAIGIATASAAAL.....EGMLFDSLPIVNFQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.\*

1: piri.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.5	9.2	262	T22599	1-acylglycerol-3-p
2	91	8.7	427	G6967	N-ethylamine ch
3	91	8.7	533	G72593	hypothetical prote
4	90	8.6	174	B83337	hypothetical prote
5	89.5	8.6	449	JC7306	extracellular prot
6	88.5	8.5	334	A71169	probable flagellin
7	88.5	8.5	1148	S72635	exo-poly-alpha-gal
8	88	8.4	378	A90280	hypothetical prote
9	86.5	8.3	284	JC5221	stomatin - mouse
10	86.5	8.3	443	F90167	conserved hypothet
11	86.5	8.3	846	S59262	hypothetical prote
12	86	8.3	1258	JQ0188	ice nucleation pro
13	85.5	8.2	329	1MBXEB	immunogenic protei
14	85.5	8.2	329	AF3351	3ik immunogenic pr
15	84.5	8.1	223	G74812	ABC-type transport
16	84.5	8.1	739	B82552	phage-related prot
17	84	8.1	391	G70625	probable PPE prote
18	83.5	8.0	458	1VGBEH	glycoprotein gp13
19	83.5	8.0	458	1B46114	glycoprotein gp13
20	83.5	8.0	870	G69006	DNA-directed RNA p
21	83.5	8.0	4436	G71086	hypothetical prote
22	83	8.0	1247	1MWHND	nidogen precursor
23	83	8.0	1322	S07053	ice nucleation pro
24	82.5	7.9	243	S58494	auxin-induced prot
25	82.5	7.9	236	D7052	homoserine kinase
26	82.5	7.9	402	G90524	conserved hypothet
27	82	7.9	147	D69008	conserved hypothet
28	82	7.9	350	A48421	ornithine transcar
29	82	7.9	1942	B71426	hypothetical prote

30 81.5 7.8 493 2 C71873 septum formation p  
31 81.5 7.8 557 2 JC5229 laccase (EC 1.10.3  
32 81.5 7.8 663 2 S51865 RGPI protein - yea  
33 81.5 7.8 712 2 A96023 probable methylmal  
34 81.5 7.8 862 2 AC1214 fibrinogen-binding  
35 81.5 7.8 1554 2 C72647 hypothetical prote  
36 81.5 7.8 2044 2 AB1180 probable peptidogl  
37 81 7.8 253 2 S76719 hypothetical prote  
38 81 7.8 388 2 E87349 HlyD family secret  
39 81 7.8 433 2 S77340 hypothetical prote  
40 81 7.8 494 2 C70667 hypothetical prote  
41 81 7.8 611 2 E90543 glucose inhibited  
42 81 7.8 660 2 E90595 hypothetical prote  
43 80.5 7.7 161 2 S15268 fima protein - Dic  
44 80.5 7.7 341 2 E83737 C4-dicarboxylate t  
45 80.5 7.7 598 2 A64594 site-specific DNA-

#### ALIGNMENTS

##### RESULT 1

T22599

1-acylglycerol-3-phosphate O-acyltransferase (EC 2.3.1.51) homolog F59P4.4 - Caenorhabd  
C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 06-Oct-2000

C:Accession: T22599; T23012

R:Dobson, R.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19587

A:Accession: T22599

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-262 <WIL>

A:Cross-references: EMBL:Z81089; PIDN:CAB03139.1; GSPDB:GN00028; CBSP:F59P4.4

A:Experimental source: clone F53H4

R:Harris, B.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19652

A:Accession: T23012

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-262 <W12>

A:Cross-references: EMBL:Z81095; PIDN:CAB03160.1; GSPDB:GN00028; CBSP:F59P4.4

A:Experimental source: clone F59F4

C:Genetics:

A:Gene: CBSP:F59P4.4

A:Map position: X

A:Introns: 57/2; 102/1; 160/3; 192/3

C:Superfamily: mouse 1-acylglycerol-3-phosphate O-acyltransferase

C:Keywords: acyltransferase; coenzyme A

Query Match 9.2%; Score 95.5; DB 2; Length 262;

Best Local Similarity 24.2%; Pred. No. 1.9;

Matches 58; Conservative 27; Mismatches 96; Indels 59; Gaps 9;

QY 4 TTTAAGIATASAAALA-----LLAGPATT---QSPINSPYATG 38  
DB 2 TFLAIFVIAVLLLLAQLFVIGYFIRAVYFGMCLITGGFLGLASIPFGKSPNNHFMFK 61  
QY 39 TAQVSEPDVESHLSG-----ITPAAGAQSDDIGVAIVIKQVNDVKLVTL 88  
DB 62 IQAMTWPGVAFELNSEILHDKKPYIIANHQSAVDVLGMSFAMPVDCV--VMLKSSL 119  
QY 89 RNAEQLPKYFKLQIQITSGYETNSFALGNFSETKAVISLDNPSAVIVLDKEDIAVLYPD 148  
DB 120 -----KYLPGNLCAYLDCSVYINRFSKALKATVDTTLHEIVTKKRWI--YPE 168  
QY 149 KTCYNTSWTSGPEPKIIVYNETK-----PVALINFKAFYEAK-----GMLFDSLPIV 198  
DB 169 GTRNAEPKL-LPFFKGAFILAKQAKTIPVCFVSSHKKFYSHAERKLTSGNCIIDLPEV 227

Db 8 VSAAAAALLAVAPVAATAMPVNAATTVTSTTTNNKPTVDLSGAGSVSESKDVTNVTSPST 67  
Qy 60 -----AGAGSDDIGYAIWIKQDVNDVKLVKTLENAEQLKPYF-----KYLQ 100

Db 68 LLSAAGKIPATLQSGTEASLNGTSVADVADAKVTLTDGKGVAVSYDKNTLTKLSD 127  
 QY 103 IQITSGYETNSTALGNFSTKA----VISLNPASVIV----LDKEDIAVLVPDKTYGYN 154  
 Db 128 VRAGDDYTWTLSGVG-FSGKANAGKTLFPLKEGTVVEGANYNKDDHKVTL-DQYGNVS 185  
 QY 155 TSIWVGPGEPKIIVNETKPVAILNPKAYEAKGML 191  
 Db 186 GLKFFV---TSKVKAYDSANTNAV----SFYDAKSGIV 215  
 RESULT 6  
 A71169  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
 R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Sekin  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kuchida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
 A:Reference number: A71000; PMID:98344137; PMID:9679194  
 A:Accession: A71169  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-334 <RAW>  
 A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29638.1; PID:g3256955  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH0549  
 C:Superfamily: archaeal flagellin  
 Query Match 8.5%; Score 88.5; DB 2; Length 334;  
 Best Local Similarity 22.1%; Pred. No. 9.6;  
 Matches 42; Conservative 30; Mismatches 63; Indels 55; Gaps 8;  
 QY 8 IAGIIASAAALALL-----AGFATTQSPINSPFYAT---GTAQAVSEPIDVES 51  
 Db 15 IAWLVAAVAVALINISGVYLOQKSQATGRTQTEVASGIKIVRVVGKADSATNPTVIOE 74  
 QY 52 HLGSIITPAAGQSGDDIGYALVWIKQDVNDVVKLVTRNAEOLKPKYKYLQITSGYET 111  
 Db 75 LAVYITPNAGSSG-----IDLTKVRLTISDG-QKQAIKPY----- 108  
 QY 112 NSTALGNFSETKAVIS--LDNPSAVIVLDKEDIAVLVPDKTYGYNNTSGIWPGEPKIIVY 169  
 Db 109 ---RVGNSANELYFLAELMQLNGATIKFDNGSVQV-YFDPTDWTISAA-----PTVID 157  
 QY 170 NETKPVAILN 179  
 Db 158 TTNKVBIVN 167  
 RESULT 7  
 S72635  
 C:Species: Thermoanaerobacterium thermosulfurigenes  
 C:Date: 29-Jul-1997 #sequence\_revision 29-Jul-1997 #text\_change 15-Oct-1999  
 C:Accession: S72635; S72621  
 R:Matuschek, M.; Sahm, K.; Bahl, H.  
 submitted to the EMBL Data Library, March 1996  
 A:Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes B  
 A:Reference number: S72635  
 A:Accession: S72635  
 A:Molecule type: DNA  
 A:Residues: 1-1148 <MATU>  
 A:Cross-references: EMBL:U0951; NID:g1542972; PIDN:AAH08040.1; PID:g1255234  
 A:Experimental source: strain EM1  
 R:Matuschek, M.; Sahm, K.; Zibat, A.; Bahl, H.  
 Mol. Gen. Genet. 252, 493-496, 1996  
 A:Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes EM1 tha

A:Reference number: S72621; PMID:97033555; PMID:8879252  
 A:Accession: S72621  
 A:Molecule type: DNA  
 A:Residues: 729-1148 <NAM>  
 A:Cross-references: EMBL:U0951  
 A:Experimental source: EM1  
 C:Genetics:  
 A:Gene: P91A  
 C:Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;  
 a xylanase A cellulose-binding repeat homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-30/Domain: signal sequence #status predicted <SIG>  
 F:31-1148/Product: exo-poly-alpha-galacturonosidase #status predicted <MAT>  
 F:969-1022/Domain: S-layer repeat homology <SLR1>  
 F:1028-1081/Domain: S-layer repeat homology <SLR2>  
 F:1092-1144/Domain: S-layer repeat homology <SLR3>  
 Query Match 8.5%; Score 88.5; DB 2; Length 1148;  
 Best Local Similarity 19.1%; Pred. No. 46;  
 Matches 37; Conservative 38; Mismatches 82; Indels 37; Gaps 6;  
 QY 20 LLAGFATTQSPINSPFYATGTAQAVSEPIDVESHLGSIITPAAGQSGDDIGYALVWIK-DQ 78  
 Db 14 LLAFLMVFTMPMSKAFADTTSSGPAAPVDQLVLSGVT-----DTGFTLVVHKPDN 64  
 QY 79 VNDV-KLVTRNAEOLKPKYKYLQITSGY-----ETNSTALGNF--SE 121  
 Db 65 YSDITDYKITSVDSVYQYVYASQNTVASQYIKQFYDNNVGDLDKDDNGNTVMSAYKISM 124  
 QY 122 TRAVISLDNPSAVIVLDKEDIAVLVPDKTYGYNNTSGIWPGEPKIIVNETKPVAILNFX 181  
 Db 125 HSPVLGKLXPTLYTTIQVQSDANKYTSPTVITQTSTPSE-----NIINVE 174  
 QY 182 AFYEAKEGMLFDGL 195  
 Db 175 STGAVGVDGLADDV 188  
 RESULT 8  
 A90280  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: A90280  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awauez, M.J.; Chan  
 Jung, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
 arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: A90280  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-378 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13814452; PIDN:AAK41496.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SSO1258  
 Query Match 8.4%; Score 88; DB 2; Length 378;  
 Best Local Similarity 22.5%; Pred. No. 12;  
 Matches 47; Conservative 31; Mismatches 67; Indels 64; Gaps 10;  
 QY 19 ALLAGFATTQSPINSPFYATGTAQAV-----SEPIDVESHLGSIITPAAG---AOGSD 66  
 Db 43 ALVAG-----SGIINSYLSIINDIPLLSRVSEPIDVRLILEKILFSAAGNCVVGSAI 98  
 QY 67 DIGVAIVW-----IKDVND-VKL-----KVLTRNAEOLKPKYKY 100  
 Db 99 SAVEMALWSLKARKSNVELKLCCKIRDSVKVYASFPFGKIDDDVLIAIRKSLERGFDL 158  
 QY 101 LQIQTSGYETNSTALG-----NFSETKAVISLDNPSAVIVLDKEDIAVLVPDKT--- 150  
 Db 159 VKL-----HQSPSTVLAAVKAIKENYKVKIADLNSP-----FDNLDAKEFVDKVKHY 208

QY	51	SHLGSIPTAAGAGQGGDDICGYAIV-----WIK-----DQV 79
Db	145	GDRGIIGSIAGLGMSGDYTYELITYRKKNWLKMSVKRVKDESTPLTFANYDYI 204
QY	80	NDVKLKV-----LRNAEOLXPYKYLQ-----IQITSGVETNSTA-----L 116
Db	205	NDTFL-ITPHGTDPILYGIRGA-SIOHLKAMELIESNEDIDFFAIFKINQSTDIHQKI 262
QY	117	GN--FSETRKAVISLDNPSAVIVLDKEDIAVLYPD-----KTGYTNTSIWVPGEPDKI 166
Db	263	GNRFYQETKKVGVQVK--VRLLEGGDVIVETTDNILLFVYKETGELNSAAKLLKKGDEI 319
QY	167	IVYNETKP 174
Db	320	VAYGAVKP 327

RESULT 11  
 SS9262  
 hypothetical protein YNL126w - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: hypothetical protein JTA846; hypothetical protein NI222; hypothetical  
 C;Species: Saccharomyces cerevisiae  
 C;date: 08-Jul-1995 #sequence\_revision 16-Feb-1996 #text\_change 21-Jul-2000  
 C;Accession: S59262; S55157; S63067; S63071; S67329  
 R;Mallet, L.; Busserreau, F.; Jacquet, M.  
 Yeast 11, 1195-1209, 1995  
 A;title: A 43.5 kb segment of yeast chromosome XIV, which contains MFA2, MEP2, CAP/SRV2  
 A;Reference number: S59241; MUID:96109932; PMID:8619318  
 A;Accession: S59262  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-846 <MAW>  
 A;Cross-references: EMBL:Z46843; NID:9861113; PIDN:CAA86899.1; PID:9861114  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1994  
 R;Mallet, L.; Busserreau, F.; Jacquet, M.  
 submitted to the EMBL Data Library, November 1994  
 A;Description: A 43.5 kb fragment of the chromosome XIV.  
 A;Reference number: S55136  
 A;Accession: S55157  
 A;Molecule type: DNA  
 A;Residues: 1-730, 'S', 732-846 <MAW>  
 A;Cross-references: EMBL:Z46843  
 R;De Antoni, A.; D'Angelo, M.; Dal Pero, F.; Sartorello, F.; Pandolfo, D.; Lanfranchi,  
 submitted to the Protein Sequence Database, April 1996  
 A;Reference number: S63047  
 A;Accession: S63067  
 A;Molecule type: DNA  
 A;Residues: 1-846 <DEA>  
 A;Cross-references: EMBL:Z71402; NID:g1302059; PID:g1302060; MIPS:YNL126w  
 A;Experimental source: strain S288C  
 R;Mallet, L.; Busserreau, F.; Jacquet, M.  
 submitted to the Protein Sequence Database, April 1996  
 A;Reference number: S63069  
 A;Accession: S63071  
 A;Molecule type: DNA  
 A;Residues: 1-846 <MAF>  
 A;Cross-references: EMBL:Z71402; NID:g1302059; PID:g1302060; MIPS:YNL126w  
 A;Experimental source: strain S288C  
 R;D'Antoni, A.; d'Angelo, M.; dal Pero, F.; Sartorello, F.; Pandolfo, D.; Lanfranchi, G  
 submitted to the EMBL Data Library, February 1996  
 A;Description: The DNA sequence of cosmid 14-13b from chromosome XIV of Saccharomyces  
 A;Reference number: S67327  
 A;Accession: S67329  
 A;Molecule type: DNA  
 A;Residues: 1-846 <DAN>  
 A;Cross-references: EMBL:Z69382; NID:g1183941; PID:g1183944  
 C;Genetics:  
 A;Gene: SGD:SPC98  
 A;Cross-references: SGD:S0005070; MIPS:YNL126w  
 A;Map position: 14L  
 C;Superfamily: Saccharomyces cerevisiae hypothetical protein YNL126w

Query Match 8.3%; Score 86.5; DB 2; Length 846;



```
QY      82 VKLKVTLR----- 89
DB      118 LRLATLYPETIHIVARKDANIKSVADLKGKRVSLDEPGSGTIVDARIIVLEAYGLTEDDI 177
QY      90 NAEOLKP-----YFKYLIQITSGYETNSTALGNFSETKAVISLNDPSAVI 135
DB      178 KAEHLKPGAGERLKGALDAYF-----FVGGYPTGAISELAINSGISLVPISSGPEADK 231
QY      136 VLDK-----EDI--AVLYPDKTGYTNTSI---WVPG--EPDKIIVNETK 173
DB      232 ILEKYSFFSKDVVPAGAYKDVAEPTLAVAAQWVTSKQPDLLI--YNITK 280

RESULT 15
S74812
ABC-type transport protein sl11623 - Synecchocystis sp. (strain PCC 6803)
N;Alternate names: protein sl11623
C;Species: Synecchocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
C;Accession: S74812
R;Kaneko, T.; Sato, S.; Kori, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimizu, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
S
A;Reference number: S74322; MUID: 97061201; PMID: 8905231
A;Accession: S74812
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <KAN>
A;Cross-references: EMBL: D90909; GB: AB001339; NID: g1652844; PID: BAA17773.1; PID: d101850
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop; transport protein
F;22-202/Domain: ATP-binding cassette homology <ABC>
F;39-46/Region: nucleotide-binding motif A (P-loop)

Query Match      8.1%; Score 84.5; DB 2; Length 223;
Best Local Similarity 22.5%; Pred. No. 12;
Matches 42; Conservative 33; Mismatches 61; Indels 51; Gaps 9;

QY      18 LALLAGFATQSPINSFYATGTAGAVSEPI-----DVESHLSITPAAGAGSGDDIGYAI 72
DB      49 LRLAGLLTQSGI-----VKLAEPMGFVFQNDHQL--VMPTVGA----DIAFGL 93
QY      73 VWIKQVNDVKLVTLRNASQL-----KPYKYLQITSGYETNSTALGNFSETKA 124
DB      94 V--KEKLSQGVQVRVGEALQAVNLAEMARRPIY-----ALSGGQKQRIAGAIARHCS 146
QY      125 VISLNDPSAVILVDKEDIIVLYPDKTGYTN--TSIWWPG-----EPDKIIV 168
DB      147 VLLDEFTALDRDSQQELVLQVKIKQRGITALTWTHRLDELDCDGAFLLEEGKIVA 206
QY      169 YNETKV 175
DB      207 QGDPKPL 213
```

Search completed: April 6, 2004, 12:41:49  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 12:34:02 ; Search time 17 Seconds  
(without alignments)  
634.031 Million cell updates/sec

Title: US-09-997-807-2

Perfect score: 1042

Sequence: 1 VKYTLAIGIISAAALAL.....EGMLFDSLPIYFQVLQVG 207

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.5	9.2	262	1	PLC1 CAEEL
2	91	8.7	427	1	YF05 METH
3	88.5	8.5	334	1	FLA3 PYRHO
4	86.5	8.3	284	1	STOM MOUSE
5	86.5	8.3	846	1	SP98 YEAST
6	86	8.3	1258	1	ICEN ERWHE
7	85.5	8.2	329	1	BCSP BRUME
8	83.5	8.0	468	1	VGLC HSEB
9	83.5	8.0	870	1	RPAL METH
10	83	8.0	1247	1	NIDO HUMAN
11	83	8.0	1322	1	ICEA PANAN
12	82.5	7.9	243	1	AXI7 ARATH
13	82.5	7.9	296	1	KHSE CLOAB
14	82.5	7.9	633	1	GIA2 FUSNN
15	82.5	7.9	1122	1	DSG2 MOUSE
16	82	7.9	350	1	OTC RANCA
17	81.5	7.8	417	1	SYS LEPIN
18	81.5	7.8	493	1	FTS4 HELPJ
19	81.5	7.8	663	1	RGPI YEAST
20	81.5	7.8	712	1	MUTB RHIME
21	81	7.8	611	1	GIDA MYCPN
22	80.5	7.7	161	1	FMAX BACNO
23	80.5	7.7	2104	1	MYS3 SCHPO
24	80	7.7	196	1	YB17 METJA
25	80	7.7	537	1	CH61 COREF
26	80	7.7	612	1	GIDA MYCPN
27	80	7.7	880	1	LYTD BACSU
28	79.5	7.6	158	1	FMAB BACNO
29	79.5	7.6	538	1	CAT1 CLOKL
30	79.5	7.6	749	1	VP4 ROTGA
31	79.5	7.6	1196	1	XPG XENLA
32	79.5	7.6	1461	1	NEOI HUMAN
33	79.5	7.6	2432	1	Y43R IRV6

34 79 7.6 407 1 ADRM RAT  
35 79 7.6 1180 1 C4AA BACTI  
36 78 7.5 407 1 ADRM MOUSE  
37 78 7.5 455 1 DEQQ ECOLI  
38 78 7.5 512 1 V233 FOWPV  
39 77.5 7.4 375 1 YKUS CAEEL  
40 77.5 7.4 537 1 YIV9 YEAST  
41 77.5 7.4 817 1 YG4B YEAST  
42 77.5 7.4 865 1 RPAI METTW  
43 77.5 7.4 1788 1 POLN SOUV3  
44 77 7.4 382 1 Y077 METJA  
45 77 7.4 404 1 SVY SYN3

## ALIGNMENTS

### RESULT 1

ID PLC1 CAEEL STANDARD; PRT; 262 AA.  
AC Q93841; Q93783;  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative 1-acyl-sn-glycerol-3-phosphate acyltransferase F53F4.4  
DE (EC 2.3.1.51) (1-AGP acyltransferase) (1-AGPAT) (lysophosphatidic acid acyltransferase) (LPAAT).  
DE F59F4.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Harris B., Dobson R.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
CC -! FUNCTION: Converts lysophosphatidic acid (LPA) into phosphatidic acid by incorporating acyl moiety at the 2 position (By similarity).  
CC -! CATALYTIC ACTIVITY: Acyl-CoA + 1-acyl-sn-glycerol 3-phosphate = CoA + 1,2-diacyl-sn-glycerol 3-phosphate.  
CC -! PATHWAY: De novo phospholipid biosynthesis; second step.  
CC -! SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -! SIMILARITY: Belongs to the 1-acyl-sn-glycerol-3-phosphate acyltransferase family.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL; Z81095; CAB03160.1;  
CC EMBL; Z81083; CAB03160.1; JOINED.  
CC EMBL; Z81089; CAB03139.1;  
CC EMBL; Z81095; CAB03139.1; JOINED.  
CC PIR; T22599; T22599.  
CC WormPep; F59F4.4; CE11552.  
CC InterPro; IPR002123; Acyltransferase.  
CC InterPro; IPR004552; AGP acyltrn.  
CC Pfam; PF01553; Acyltransferase; 1.  
CC SMART; SM00563; PlsC; 1.  
CC TIGRFAMs; TIGR00530; AGP acyltrn; 1.  
CC Hypothetical protein; Phospholipid biosynthesis; Transferase;  
KW Acyltransferase; Transmembrane.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 29 49 POTENTIAL.  
FT TRANSMEM 89 109 POTENTIAL.  
SQ SEQUENCE 262 AA; 29638 MW; 0361FE6C9710593E CRC64;



```
Query Match          9.2%; Score 95.5; DB 1; Length 262;
Best Local Similarity 24.2%; Pred. No. 0.72;
Matches 58; Conservative 27; Mismatches 96; Indels 59; Gaps 9;

QY 4 TTALAAGIASAALA-----LLAGPATT---OSPLNSFYATG 38
DB 2 TFLAIFVIAVILLAAQLPVIGFIRAVYFGMCLIIIGFGLGLASIPFGSPNHFEMFX 61
QY 39 TAQAVSEPIDVESHLS-----ITPAAGQSGDDIGYAIWTKQVNDVKLVKTL 88
DB 62 IFQAMTWPMGVRFELRNSEILHDKKPYIIITANHQSALDVLGMSFAPVDCV---VMLKSSL 119
QY 89 RNAEQLKPYKYLOIQTSGVETNSTALGNFSEFKAVISLDNPSAVILDKEDIAVLYPD 148
DB 120 -----KYLPGFNLCAVLCDSVYINRPSKALKUTVDTLHEIVTKRKWMI-YPE 168
QY 149 KTGYNSTSIWPGEPDKIIIVNETK-----PVAILNFKAPEAKE-----GMLFDSLPVI 198
DB 169 GTRNAEPFL-LPFFKGAFILAKQAKIPIVPCVFSHSHFFYSHABKRLTSGNIGIIDLPEV 227

RESULT 2
YF05 METH
ID YF05 METH STANDARD; PRT; 427 AA.
AC 027549;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MTH1505.
GN MTH1505.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook K., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- SIMILARITY: Belongs to the ATZ/TRZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000910; AAB85980.1; -
CC PIR; G69067; G69067.
CC InterPro; IPR006680; Amidohydro 1.
CC Pfam; PF01979; Amidohydro_1; 1.
CC ProDom; PD000518; Urease; 1.
CC Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 427 AA; 46441 MW; 439527DD142182CD CRC64;

Query Match          8.7%; Score 91; DB 1; Length 427;
Best Local Similarity 22.6%; Pred. No. 3.1;
Matches 40; Conservative 29; Mismatches 80; Indels 28; Gaps 6;

QY 22 AGFATTQSPNSFYATGTAQAVSEPIDVESHLSITPA-----AQAQSGDDIGYAIWIK 76

Db 197 AALADKNDLMIHIVHSETENEYSE-----VSRSHGMTPEVLDVGVGLPRTVAACHCVWLK 252
QY 77 DQVNDV-----KUKVT-----LRNAEQLKPYKYLOIQTSGVETNSTALGN-----FSET 122
Db 253 DWEIDVLARDVKVSVSNPNKSLASGVFARLLQRGVNVSLGTGGAASNNLDFQEM 312
QY 123 KAV-----ISLDNPSAVILDKEDIAVLYPDKTGYNTSIWVPGEPDKIIIVNETKP 174
Db 313 KTSLLQKNLESDPTALPMDVFSMATLNGARALGDAGLIAEKLADIVILNTRRP 369

RESULT 3
FLA3 PYRHO
ID FLA3 PYRHO STANDARD; PRT; 334 AA.
AC OS8284;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable flagellin PH0549.
GN PH0549.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3; PubMed=9679194;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RA "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- FUNCTION: Flagellin is the subunit protein which polymerizes to
CC form the filaments of archaeal flagella (By similarity).
CC -!- SIMILARITY: Belongs to the archaeal flagellin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP000002; BAA29638.1; -
CC PIR; A71169; A71169.
CC InterPro; IPR002774; Arch_flagellin.
CC Pfam; PF01917; Arch_flagellin; 1.
CC Kw Flagellum; Multigene family; Complete proteome.
SQ SEQUENCE 334 AA; 35438 MW; 9CEAAA6B07C92296 CRC64;

Query Match          8.5%; Score 88.5; DB 1; Length 334;
Best Local Similarity 22.1%; Pred. No. 3.7;
Matches 42; Conservative 30; Mismatches 63; Indels 55; Gaps 8;

QY 8 IAGIASAALALL-----AGFATTQSPNSFYAT---GTAQAVSEPIDVES 51
Db 15 IAVLVAAVAALVINTSGVLQKQSQATGRTQTEVASGKVTITVGVGKADSATNPTVIOE 74
QY 52 HLGSIITPAAGQSGDDIGYAIWTKQVNDVKLVRLNABQLKPYKYLOIQTSGVET 111
Db 75 LAVYITFNAGSSG-----IDLTKVITLSDG-QKQAIKFI 108
QY 112 NSTALGNFSETKAVIS--LDNPSAVILDKEDIAVLYPDKTGYNTSIWVPGEPDKIIIV 169
Db 109 ---RVGSANELYFLAELMQLGATIKFDNGSVQV-YFDPDWTISAA-----PTVIID 157
QY 170 NETKPAVLN 179
```

Db 158 TTNKVIEIVN 167

| : | :

RESULT 4

STOM\_MOUSE STANDARD; PRT; 284 AA.

ID STOM\_MOUSE Q60744; Q62455;

AC P54116; Q60744; Q62455;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Erythrocyte band 7 integral membrane protein (Stomatin) (Protein 7.2b).

DE STOM OR EPB72 OR EPB7.2.

GN Mus musculus (Mouse).

OS Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Eukaryota; Metazoa; Chordata; Vertebrata; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RN [1]

RC STRAIN=BALB/c; TISSUE=Bone marrow;

RC MEDLINE=97080556; PubMed=8921901;

RX Schlegel W., Unfried I., Prohaska R.;

RA "Cloning and analysis of a cDNA encoding the BALB/c murine erythrocyte band 7 integral membrane protein.";

RT Gene 178:115-118(1996).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J;

RC Gallagher P.G., Roman M., Lieman J.H., Ward D.C.;

RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RC SEQUENCE FROM N.A.

RP MEDLINE=96374834; PubMed=8786142;

RX Gallagher P.G., Turetsky T., Mentzer W.C.;

RA "Genomic organization and 5'-flanking DNA sequence of the murine stomatin gene (Epb72).";

RT Genomics 34:410-412(1996).

RL [-] FUNCTION: Thought to regulate cation conductance.

CC [-] SUBCELLULAR LOCATION: EXPOSED ON THE CYTOPLASMIC SURFACE OF THE MEMBRANE.

CC [-] SIMILARITY: Belongs to the band 7 / mec-2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

DR EMBL; X91043; CAA62503.1; .

DR EMBL; U17297; AAN75024.1; .

DR EMBL; U50999; AAB18857.1; ALT SEQ.

DR EMBL; U50993; AAB18857.1; JOINED.

DR EMBL; U50994; AAB18857.1; JOINED.

DR EMBL; U50995; AAB18857.1; JOINED.

DR EMBL; U50996; AAB18857.1; JOINED.

DR EMBL; U50997; AAB18857.1; JOINED.

DR EMBL; U50998; AAB18857.1; JOINED.

DR PIR; JC5221; JC5221.

DR MGD; MG1:95403; Epb7.2.

DR InterPro; IPR001107; Band 7.

DR InterPro; IPR001972; Stomat.in.

DR Pfam; PF01145; Band\_7; 1.

DR PRINTS; PR00721; STOMATIN.

DR SMART; SM00244; PHB; 1.

DR PROSITE; PS01270; BAND 7; 1.

KW Erythrocyte; Transmembrane; Phosphorylation; Lipoprotein; Palmitate.

FT TRANSMEM 32 52

POTENTIAL.

FT DOMAIN 53 284

CYTOPLASMIC (POTENTIAL).

FT LIPID 30 30

S-palmitoyl cysteine (By similarity).

FT LIPID 87 87

S-palmitoyl cysteine (By similarity).



RA Mayfield J.E., Bricker B.J., Godfrey H., Crosby R.M., Knight D.J.,  
RA Halling S.M., Balinsky D., Tabatabai L.B.;  
RT "The cloning, expression, and nucleotide sequence of a gene coding  
RT for an immunogenic Brucella abortus protein.";  
RL Gene 63:1-9(1988).  
CC -!- MISCELLANEOUS: BRUCELLA ABORTUS IS THE CAUSATIVE AGENT FOR  
CC BRUCELLOSIS IN CATTLE AND MAN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AB009521; AAL51977.1; -.  
DR EMBL; M20404; AAA22993.1; -.  
DR PIR; AF3351; AF3351.  
DR PIR; J02079; IMKB3.  
KW Signal; Complete proteome.  
FT SIGNAL 1 28  
FT CHAIN 29 329 31 kDa IMMUNOGENIC PROTEIN.  
SQ SEQUENCE 329 AA; 34273 MW; 4EPA2330D0A2544E CRC64;  
Query Match 8.2%; Score 85.5; DB 1; Length 329;  
Best Local Similarity 19.3%; Pred. No. 6.5; Mismatches 73; Indels 133; Gaps 13;  
Matches 56; Conservative 28;  
QY 2 KYTLAAGIASAALALLAGPATTQSPINSPYATGTAQVSEPIDVESHLSITPAAG 61  
DB 6 KTRLLAAVAV---AGAILGASFAVAQAP--TFRIGTGCTAGTYPIGGLIANASGAG 60  
QY 62 AQG-----SDDIGVAIWI-----KQOVND 81  
DB 61 EKGVPGLVATVSSNGSVANAINAKSGALESGTQSDVAY---WAYNGTGLYDGKGVED 117  
QY 82 VKLKVTLR----- 89  
DB 118 LRLALLPTETHIVARKDANIKSVADLKGKRVSLDEPGSGTIVDARIVLEAVGLTDDI 177  
QY 90 NARQLKP-----YFKYLQITTSYETNSTALGNFSETKKAVISLDNPSAVI 135  
DB 178 KAEHLKPGPAGERLKDGLDAYF-----FVGGYPTGALSELAINSGISLVPSISGPEADK 231  
QY 136 VLDK-----EDI--AVLYPDKTGYNTSI---WVPG--BPDKLIIVNETK 173  
DB 232 ILEKYSFFSKDVPVPAGAYKDVETPTLIAVAQWVTSKQPDLLI--YNITK 280  
RESULT 8  
VGLC\_HSVB STANDARD; PRT; 468 AA.  
AC P12889; P36321;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Glycoprotein C precursor (Glycoprotein 13).  
GN GC OR GP13 OR 16.  
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1), and  
OS Equine herpesvirus type 1 (strain Kentucky D) (EHV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=31520, 10330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AB4p;  
RX MEDLINE=92295566; PubMed=1318606;  
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
RL "The DNA sequence of equine herpesvirus-1.";  
RN Virology 189:304-316(1992).  
RP [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=Kentucky D;  
RX MEDLINE=88275055; PubMed=2455821;  
RA Allen G.P., Coogale L.D.;  
RT "Characterization of an equine herpesvirus type 1 gene encoding a  
RT glycoprotein (gp13) with homology to herpes simplex virus  
RT glycoprotein C.";  
RL J. Virol. 62:2850-2858(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kentucky D;  
RX MEDLINE=89382761; PubMed=2550665;  
RA Guo P., Goebel S., Davis S., Perkus M.E., Languet B., Desmettre P.,  
RA Allen G., Paoletti E.;  
RT "Expression in recombinant vaccinia virus of the equine herpesvirus 1  
RT gene encoding glycoprotein gp13 and protection of immunized  
RT animals.";  
RN J. Virol. 63:4189-4198(1989).  
RL [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kentucky D;  
RX MEDLINE=93212524; PubMed=8384760;  
RA Matsumura T., Smith R.H., O'Callaghan D.J.;  
RT "DNA sequence and transcriptional analyses of the region of the  
RT equine herpesvirus type 1 Kentucky A strain genome encoding  
RT glycoprotein C.";  
RL Virology 193:910-923(1993).  
CC -!- SIMILARITY: Belongs to the herpesviruses glycoprotein C family.  
CC -!- SIMILARITY: Contains 2 immunoglobulin-like domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; L07272; AAA46078.1; -.  
DR EMBL; M86664; AAB02451.1; -.  
DR EMBL; M19966; AAA46077.1; -.  
DR EMBL; M29234; AAA46085.1; -.  
DR EMBL; S57839; AAB25944.1; -.  
DR PIR; A28149; VGBEEH.  
DR PIR; A28114; B46114.  
DR InterPro; IPR001038; Gp13\_EHV.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR001654; Marek\_A.  
DR Pfam; PF02124; Marek\_A; 1.  
DR PRINTS; PR00668; GLYCOPROTEIN.  
DR PROSITE; PS50835; IG LIKE; 2.  
KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal; Repeat.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 468 GLYCOPROTEIN C.  
FT TRANSMEM 432 451 POTENTIAL.  
FT DOMAIN 220 311 IG-LIKE 1.  
FT DOMAIN 321 416 IG-LIKE 2.  
FT CARBOHYD 45 46 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 62 62 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 269 269 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 107 107 E -> K (IN REF. 4).  
FT CONFLICT 145 145 E -> K (IN REF. 4).  
FT CONFLICT 275 275 V -> A (IN REF. 4).  
SQ SEQUENCE 468 AA; 50889 MW; 86F67AFD15AF1C89 CRC64;  
Query Match 8.0%; Score 83.5; DB 1; Length 468;  
Best Local Similarity 22.5%; Pred. No. 15;  
Matches 46; Conservative 33; Mismatches 72; Indels 53; Gaps 9;

```

QY 6 LAIAGIIASAAALLAGFAFTTOSPLNSFYATGTAQAVSEPIDVESHLSGITPAAGAGQS 65
Db 10 VAVAYLICAGAILTYASGASASSQ-----STPATPTHTTPLLTTAAGA-GS 55
QY 66 DDTGYALVWKQDVNDVKLVKLTARNARQLKPKYFKYLOIQITSGVETNSTALG-NFSETKA 124
Db 56 DNTTANG--TESTHSHETTITCTKSLISVPYKQSDVMNCT-----TSVGVNYSYRL 106
QY 125 VISLNDPSAVILDKEDIAVLDPKGTNTSIWVGPEDKIIVYNETKPVAILNPKAFY 184
Db 107 EIVLN-----QRTFSGTP-----PDDENYINHNATKQDTLLLFSTAE 145
QY 185 EAKE-----GMLFDSLPL--VIFN 200
Db 146 RKKSRRGGQLGVIPDRLPKROLFN 169

RESULT 9
RPAL METHH STANDARD; PRT; 870 AA.
AC 027125;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6).
GN RPOA1 OR MTH1051.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RC MEDLINE=98037514; PubMed=93711463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- COFACTOR: Zinc.
CC -!- SUBUNIT: M.thermoautotrophicum RNA polymerase is composed of four
CC subunits: A', A'', B' and B''.
CC -!- SIMILARITY: THE COMBINED A'+A" SUBUNITS CORRESPOND TO THE A
CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
CC EUKARYOTIC BETA' SUBUNIT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000876; AAB85542.1; -.
CC PIR; G69006; G69006.
CC InterPro; IPR000722; RNA_pol_A.
CC InterPro; IPR007080; RNA_pol_Rpb1_1.
CC InterPro; IPR007066; RNA_pol_Rpb1_3.
CC InterPro; IPR007083; RNA_pol_Rpb1_4.
CC InterPro; IPR007081; RNA_pol_Rpb1_5.
CC InterPro; IPR006592; RNA_pol_N.

```

```

DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR SMART; SM00663; RPOA_N; 1.
KW Transfrase; DNA-directed RNA polymerase; Transcription; Zinc;
KW Complete proteome; Zinc-finger.
FT ZN FING 60 103 C4-TYPE (POTENTIAL).
SQ SEQUENCE 870 AA; 98077 MW; 05BB2C280CFD211 CRC64;

Query Match 8.0%; Score 83.5; DB 1; Length 870;
Best Local Similarity 21.7%; Pred.No. 31;
Matches 46; Conservative 32; Mismatches 73; Indels 61; Gaps 9;

QY 46 PIDVESHLSGITPAAGAGQSDDIGYAIVWIKQDVNDVKLVKLTARNARQL--KPYFKYLQ 102
Db 212 PVTVRP---SITLETGERSEDDLTHTKLVDIL-RINQ-RUKENMEAGAPQLIVEDLWELLQ 266
QY 103 IQITSGVETNSTALG-----NFSETKAV----- 125
Db 267 YHVTTFVDNEASGVPPARHRSGRPLKTLAQLKKEGRFESNLGKRVNFSARTVISPDP 326
QY 126 -ISLNDPSAVILDKEDIAVLDPKGTNTSIWVGPED-----KIIVYN 170
Db 327 NISINEVGVPETIAREVTVPVYVTEWNIDRMRYIENGPDVHFGANVIRPDGKRIYN 386
QY 171 ETRKPVAILNPKAFYEAKEGMLFDSLPLVIFNFQ 202
Db 387 ETRKPVILENLKPGY-IVERHLKDGDIVLFPNRQ 417

RESULT 10
NIDO HUMAN STANDARD; PRT; 1247 AA.
AC P14543; Q14942;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nidogen precursor (Entactin).
GN NID.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90091745; PubMed=2574658;
RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
RA Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
RT "Human nidogen: complete amino acid sequence and structural domains
RL deduced from cDNAs, and evidence for polymorphism of the gene.";
RNA 8:581-594(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=9604428; PubMed=7557988;
RA Zimmermann K., Hoischen S., Hafner M., Nischt R.;
RT "Genomic sequences and structural organization of the human nidogen
RL gene (NID).";
RNA Genomics 27:245-250(1995).
RN [3]
RP SEQUENCE OF 567-1247 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89270475; PubMed=2471408;
RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Weil D.,
RA Timpl R., Chu M.L., Uitto J.;
RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the
RL gene to chromosome 1q43.";
RNA Am. J. Hum. Genet. 44:876-885(1989).
CC -!- FUNCTION: Sulfated glycoprotein which is widely distributed in
CC basement membranes and that is tightly associated with laminin.
CC Also binds to collagen IV. It probably has a role in cell-
CC extracellular matrix interactions.

```

```

CC CC -1- SUBUNIT: Interacts with FBLN1 (By similarity).
CC CC -1- SUBCELLULAR LOCATION: Basement membranes.
CC CC -1- PTM: N- and O-glycosylated.
CC CC -1- SIMILARITY: Contains 6 EGF-like domains.
CC CC -1- SIMILARITY: Contains 1 thyroglobulin type-I domain.
CC CC -1- SIMILARITY: Contains 5 LDL-receptor YWTD domains.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M30269; AAA59932.1; -
CC CC EMBL; X82245; CAA57709.1; -
CC CC EMBL; X84819; CAA57709.1; JOINED.
CC CC EMBL; X84820; CAA57709.1; JOINED.
CC CC EMBL; X84821; CAA57709.1; JOINED.
CC CC EMBL; X84822; CAA57709.1; JOINED.
CC CC EMBL; X84823; CAA57709.1; JOINED.
CC CC EMBL; X84824; CAA57709.1; JOINED.
CC CC EMBL; X84825; CAA57709.1; JOINED.
CC CC EMBL; X84826; CAA57709.1; JOINED.
CC CC EMBL; X84827; CAA57709.1; JOINED.
CC CC EMBL; X84828; CAA57709.1; JOINED.
CC CC EMBL; X84829; CAA57709.1; JOINED.
CC CC EMBL; X84830; CAA57709.1; JOINED.
CC CC EMBL; X84831; CAA57709.1; JOINED.
CC CC EMBL; X84832; CAA57709.1; JOINED.
CC CC EMBL; X84833; CAA57709.1; JOINED.
CC CC EMBL; X84834; CAA57709.1; JOINED.
CC CC EMBL; X84835; CAA57709.1; JOINED.
CC CC EMBL; X84836; CAA57709.1; JOINED.
CC CC EMBL; X84837; CAA57709.1; JOINED.
CC CC EMBL; M27445; AAA57261.1; -
CC CC PIR; A33322; MMHUND.
CC CC PDB; 1NDX; 29-DEC-99.
CC CC Genew; HGNC:7821; NID.
CC CC MIM; 131390; -
CC CC InterPro; IPR000152; Asx_hydroxyl_5.
CC CC InterPro; IPR001881; EGF_Ca.
CC CC InterPro; IPR006209; EGF_Like.
CC CC InterPro; IPR006605; GZF_Like.
CC CC InterPro; IPR009017; GPF_Like.
CC CC InterPro; IPR000033; Ldl_receptor_rep.
CC CC InterPro; IPR003886; Nidogen_ext.
CC CC InterPro; IPR00716; Thyroglobulin_1.
CC CC Pfam; PF00008; EGF; 6.
CC CC Pfam; PF00058; ldl_recept_b; 3.
CC CC Pfam; PF00086; thyroglobulin_1; 1.
CC CC SMART; SM00179; EGF_CA; 2.
CC CC SMART; SM00682; GZF; 1.
CC CC SMART; SM00135; LY; 5.
CC CC SMART; SM00539; NIDO; 1.
CC CC SMART; SM00211; TY; 1.
CC CC PROSITE; PS00010; ASX_HYDROXYL; 3.
CC CC PROSITE; PS00022; EGF_1; 1.
CC CC PROSITE; PS01186; EGF_2; 5.
CC CC PROSITE; PS00026; EGF_3; 5.
CC CC PROSITE; PS01187; EGF_CA; 2.
CC CC PROSITE; PS00484; THYROGLOBULIN_1; 1.
CC CC Basalment membrane; Extracellular matrix; Glycoprotein; Sulfation;
CC CC Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion;
CC CC 3D-structure.
CC CC SIGNAL
CC CC CHAIN 1 28
CC CC CHAIN 29 1247
CC CC FT CHAIN 29 669
CC CC FT DOMAIN 29 917
CC CC FT DOMAIN 670 917
CC CC FT DOMAIN 918 1247
CC CC FT DOMAIN 386 426
CC CC FT DOMAIN 668 709
CC CC FT NIDOGN.
CC CC FT I (LARGER GLOBULAR DOMAIN).
CC CC FT II (CYSTEINE-RICH).
CC CC FT III (SMALLER GLOBULAR DOMAIN).
CC CC FT EGF-LIKE 1.
CC CC FT EGF-LIKE 2.

```

```

FT DOMAIN 710 751 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 758 801 EGF-LIKE 4.
FT DOMAIN 802 840 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 872 919 THYROGLOBULIN TYPE-I.
FT DOMAIN 989 1030 LDL-RECEPTOR YWTD MOTIF 1.
FT DOMAIN 1032 1073 LDL-RECEPTOR YWTD MOTIF 2.
FT DOMAIN 1075 1118 LDL-RECEPTOR YWTD MOTIF 3.
FT DOMAIN 1124 1163 LDL-RECEPTOR YWTD MOTIF 4.
FT DOMAIN 1208 1244 EGF-LIKE 6.
FT MOD_RES 289 289 SULFATION (POTENTIAL).
FT MOD_RES 296 296 SULFATION (POTENTIAL).
FT DISULFID 672 685 BY SIMILARITY.
FT DISULFID 679 695 BY SIMILARITY.
FT DISULFID 697 708 BY SIMILARITY.
FT DISULFID 714 727 BY SIMILARITY.
FT DISULFID 721 736 BY SIMILARITY.
FT DISULFID 738 750 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 769 787 BY SIMILARITY.
FT DISULFID 789 800 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 839 BY SIMILARITY.
FT DISULFID 1212 1223 BY SIMILARITY.
FT DISULFID 1219 1232 BY SIMILARITY.
FT DISULFID 1234 1243 BY SIMILARITY.
FT SITE 702 704 CELL ATTACHMENT SITE.
FT CARBOHYD 1137 1137 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 33 34 EL -> SS (IN REF. 2).
FT CONFLICT 37 42 FGPGQG -> SAPDR (IN REF. 2).
FT CONFLICT 1115 1115 H -> T (IN REF. 3).
SQ SEQUENCE 1247 AA; 136488 MW; 4681B5B3CEC1575B CRC64;

Query Match 8.0%; Score 83; DB 1; Length 1247;
Best Local Similarity 23.7%; Pred. No. 52;
Matches 47; Conservative 23; Mismatches 72; Indels 56; Gaps 9;

QY 28 QSPINSFYATGTA-QAVSRPIDVESHGSIPTAAGA-----QGSDDIGYALVMIKQ 78
DB 68 RSDIDAVYTTNGIATSEPPAKESHPLPPTFGAVAPLADLDTDDGUGK--VYIRE- 124

QY 79 VNDVKLVTLRANAEOLKPKFYKLIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLD 138
DB 125 --DLSPSITQRAAECHVRGPFESFQ-----PSSAVVVT 156

QY 139 KEDIAVLY-----PDKTYTNTSIWPGEPDK-----IIVNETKPVALLNFKAFYEAKG 189
DB 157 WESVAPYQGFSPRDPDQKRGNTFQAVLASSDSSSYAIFLYPEDG----LQFHTTFSKKN 212

QY 190 MLFDSLPIVFNFQVLQVG 207
DB 213 ---NQVPVVVAFSQSGVG 227

RESULT 11
ICEA PANAN
ID ICEA PANAN STANDARD; PRT; 1322 AA.
AC P20459;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaA.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
CX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice

```

```

RT nucleation activity."
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17316; CA35194.1; -.
CC PIR; S07053; S07053.
CC HSSP; P06620; 11NA.
CC InterPro; IPR000258; Ice nucleatn.
CC Pfam; PF00818; Ice nucleation; 69.
CC PRINTS; PR00327; ICNNUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 49.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1322 AA; 131094 MW; 89B0EB24A837039 CRC64;

Query Match 8.0%; Score 83; DB 1; Length 1322;
Best Local Similarity 23.2%; Pred. No. 56;
Matches 46; Conservative 27; Mismatches 65; Indels 60; Gaps 11;

QY 3 YTTIAIAGIIASAALALLAGFATTQSP-LNSFYAT--GTAQAVSEPIDVESHGSLTPA 59
Db 644 YGSTSTAG-----ADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQEGSDLTAGVGS-TST 697
QY 60 AGAQGSDDIGYAIWIKDQNDVVKVLTNRNAEQLEPKFKYL-----QIQITS 107
Db 698 AGADSSLIAGY-----GSTQTASYHSSLTAGYGSTQTAQEGSVLIT 738
QY 108 GYETNSTA-----LGNFSETK-----AVISLDNPSAVIVLDKEDIAVLVPEDKTGYNTS 156
Db 739 GYGSTSTAGADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQERSDLT-----TGYGSTS 792
QY 157 IWVPGEPDKIIV-YNEPK 173
Db 793 --TAGADSSLIAGYGSTQ 808

RESULT 12
AX17 ARATH STANDARD; PRT; 243 AA.
ID AX17 ARATH
AC Q38825;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Auxin-responsive protein IAA7 (Indoleacetic acid-induced protein 7).
GN IAA7 OR AT3G23050 OR MXC7.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OK NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=95387393; PubMed=7658471;
RA Abel S., Nguyen M.D., Theologis A.;
RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in
RL Arabidopsis thaliana."
J. Mol. Biol. 251:533-549(1995).

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.,
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954450; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newman G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Heuan V.W., Iida K., Karnes M.,
RA Khan S., Koeseema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vayaberg M., Wallander E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome."
RL Science 302:842-846(2003).
CC -!- FUNCTION: Could act as regulator of genes responsible for
CC mediating the various auxin-induced events responsible for cell
CC growth (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- INDUCTION: By auxin.
CC -!- SIMILARITY: Belongs to the AUX/IAA family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18409; AAC49048.1; -.
CC PIR; S58494; S58494.
CC TRANSFAC; T04536; -.
CC InterPro; IPR003311; AUX_IAA.
CC Pfam; PF02309; AUX_IAA.
KW Multigene family; Nuclear protein; Translation regulation.
SQ SEQUENCE 243 AA; 26381 MW; 830EB9EA2C2B5D0 CRC64;

Query Match 7.9%; Score 82.5; DB 1; Length 243;
Best Local Similarity 25.0%; Pred. No. 8;
Matches 54; Conservative 34; Mismatches 79; Indels 49; Gaps 12;

QY 13 ASAAALALLAGFATTQSP-LNSFYATGTAQAVSEPIDVESHGSLTPAAGQSDDIGAI 72
Db 10 ATELCLCLPGGAEEVSEPAKS--AVGSKRGFSETVDMLMLNLS-----NKEGSVD----- 57
QY 73 VTIKDQNDVVKVLTNRNAEQ-----LKYFKYLOIQ--ITSG-YETNSTALG 117
Db 58 --LKNVSAVPEKKTLLKDPKPPAKQVGVGPPVRYRKNMTTQQTSSGAEESEKAG 115
QY 118 NFSETKA-----VISLDNPSAVIVLDKEDIAVL--YPDKTG-----YTNTSIWVPGEPDK 165
Db 116 NFGGAGAGAGLVKVSMDGAP---YLRKVDLKMYSQYQDLSDALAKMFSSFTMGYGAQM 172
QY 166 IIVNETKPVAILNFKAF---YFAKEG--MLFDSL 196
Db 173 IDFMNESKLMNLNLSSEYVPSYEDKGDGMWLVGDVP 208

```





[illegible]



QY 118 NFSETKAVIS---LNPNSAVILVD--KEDIAVLVPDKTGYTNTSWVGPDPKII-VVNE 171  
 Db 350 QLELEVRLOGEFAQHPSQPLVIAALKGVLSLALGDRINYNVNASLEAKGRGIRVLEVKDB 409

QY 172 T 172  
 Db 410 T 410

## RESULT 2

Q33651 ID Q33651 PRELIMINARY; PRT; 255 AA.  
 AC Q53651;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE CAMP factor precursor.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1311;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R268;  
 RX MEDLINE=95231384; PubMed=7715536;  
 RA Podbielski A.;  
 RT "Molecular characterization of the *cf6* gene encoding group B streptococcal CAMP-factor";  
 RL Med. Microbiol. Immunol. 183:239-256(1994).  
 DR EMBL; X72754; CAA51283.1; --  
 KW Signal; CAMP.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 255 POTENTIAL.  
 SQ SEQUENCE 255 AA; 28373 MW; 5061CE85C31DF994 CRC64;

Query Match 9.2%; Score 96; DB 2; Length 255;

Best Local Similarity 25.0%; Pred. No. 4.5;  
 Matches 53; Conservative 31; Mismatches 84; Indels 44; Gaps 10;

QY 5 TLAIAGIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVES-HLGSITPAAG 63  
 Db 12 TLVAGALLFSPVLEHVADQVTTTPQVHVNNVNNQAQMAQKLDQDSITLQENIKD--NVQ 69  
 QY 64 GSD---DIGYAIWIKQVNDVVKLTIR-----NARQLKPFYKYLQIT 106  
 Db 70 GTDYKPVNEAITSVE-----KLKTSLRANPETVYDLSIGSRVEALTDVIE--AITFS 121

QY 107 SGYETNSTALGN---PSETKAVISLDNPSAVI-----VLDKEDIAVLVPDKTGYT 153  
 Db 122 TQHLTNKVSQANIDMGFGITKLIVIRLDPFASVDSIKAVQNDVKALEQKVLTPDLKPTD 181

QY 154 NTSIWVGPDPKIIIVNE--TKPVAILNFKAF 183  
 Db 182 RATIYTKSKLDKEI-WNTRFTRDKKVLNVKFE 212

## RESULT 3

Q8VW91 ID Q8VW91 PRELIMINARY; PRT; 449 AA.  
 AC Q8VW91;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Proteinase.  
 GN PRT.  
 OS Lactobacillus helveticus.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=1587;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCM1003;

RA Saito Y., Hamanaka Y., Takizawa S., Benno Y.;  
 RT "Rapid Detection of *Lactobacillus helveticus* in Fecal Samples of Healthy Subjects Administered *L. helveticus* Yogurt using S-layer Gene-Targeted Primers";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB061776; BAB72066.1; --  
 DR GO; GO:0009274; C:cell wall (sensu Bacteria); IEA.  
 DR GO; GO:0030115; C:S-layer; IEA.  
 DR GO; GO:0005193; E:structural constituent of cell wall; IEA.  
 DR InterPro; IPR004903; SLAP.  
 DR Pfam; PF03217; SLAP; 1.  
 DR PRINTS; PRO1729; SURFACELAYER.  
 SQ SEQUENCE 449 AA; 47940 MW; EBA76530777D343A CRC64;

Query Match 9.1%; Score 95; DB 2; Length 449;

Best Local Similarity 24.9%; Pred. No. 12;  
 Matches 54; Conservative 31; Mismatches 86; Indels 46; Gaps 9;

QY 11 IIASAALALLAGFATTQSPINSFYATGTAQAVSEPIDVESHLGS-----ITPA-- 59  
 Db 7 IVSAAALAVAPVAATAMPVNAATTVTSTTNKPTVDLSGAGSVSESKDTVNVTPSFT 66  
 QY 60 -----AGAGSDDIGYAIWIKQVNDVVKLTIRNAEQKPFYKYLQIQ----- 104  
 Db 67 LTSAAKGIPATLQGSIEASLNGTSVTADVADVAKDVTLD-DGGKTVISYNEGKKVDNN 125  
 QY 105 ----ITSGYETNSTALG-NFSETKA-----VISLDNPSAVILVD-KEDIAVLVPDKTGYTNT 155  
 Db 126 LSAVEAGKEYTMTLSGVGFSFGKANAGKTLTFKLPKVKVNDTSNDVKVSLDQYGNATNL 185  
 QY 156 SIWVGPDPKIIIVN-ETKPVAILNFKAFYEAKGML 191  
 Db 186 -----KFVISNIKAYDSANTNAVSVFYDAKSGLV 213

## RESULT 4

Q8ON15 ID Q8ON15 PRELIMINARY; PRT; 1666 AA.  
 AC Q8ON15;  
 DT 01-JUN-2003 (T-EMBLrel. 24, Created)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Structural glycoprotein.  
 OS Yellow head virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales.  
 OX NCBI\_TaxID=96023;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22541857; PubMed=12655087;  
 RA Jitrapakdee S., Unajak S., Sittidilokratna N., Hodgson R.A.J., Cowley J.A., Walker P.J., Panyim S., Boonsaeng V.;  
 RT "Identification and analysis of gp116 and gp64 structural glycoproteins of yellow head nidovirus of *Penaeus monodon* shrimp.";  
 RL J. Gen. Virol. 84:863-873(2003).  
 DR EMBL; AF540644; AAC83987.1; --  
 SQ SEQUENCE 1666 AA; 185430 MW; C0B5223A754D6EA9A CRC64;

Query Match 9.0%; Score 94; DB 12; Length 1666;

Best Local Similarity 26.4%; Pred. No. 95;  
 Matches 51; Conservative 29; Mismatches 71; Indels 42; Gaps 12;

QY 20 LLAGFATTQSPINSFYATGTAQAVSEPID--VEHLGSITPAAGAQSDDIGYAIWIKD 77  
 Db 862 LYLGVSAASIGTYGCKGAQSIPIYHDHGHINTDLG--TPVYDS-ACDSAAATIPVVVKY 918

QY 78 Q-----VNDVVKLTIRNAEQKPFYKYLQIQITSGVETNSTALGNFSETKAVISLDNP 131  
 Db 919 NGPYSLGVPDVSCIEI---HDETL-----TCG--TNSTP--RFSICSHKLPVDPG 960

QY 132 SAVIVLDKEDIAVLVPDKTGYTNTSIWVGPDPKII-IVYNETKP-----VALNFKA 182  
 Db 961 HSVTCINSKDNKVVVKQPGY---SYIAGDPGALHSHNKHKPYTSLKQINLHFESY 1017

```

QY 183 FYEKEGMLFDSL 195
Db 1018 LYQA-VAMLFGLS 1029

RESULT 5
Q9LDC3 PRELIMINARY; PRT; 508 AA.
ID Q9LDC3
AC Q9LDC3;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Dipeptide transporter OppA homolog.
GN DPAA.
OS Ochrobactrum anthropi.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Ochrobactrum.
OX NCBI_TaxID=529;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRC SV3;
RX MEDLINE=20193627; PubMed=10727942;
RA Komeda H., Asano Y.;
RT "Gene cloning, nucleotide sequencing, and purification and
RT characterization of the D-sterespecific amino-acid amidase from
RT Ochrobactrum anthropi SV3."
RL Eur. J. Biochem. 267:2028-2035 (2000).
DR EMBL; AB026907; BAA94699.1; -.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
DR SQA 508 AA; 55661 MW; 845795A526CA74E CRC64;

Query Match 8.8%; Score 91.5; DB 2; Length 508;
Best Local Similarity 27.6%; Pred. No. 29;
Matches 40; Conservative 23; Mismatches 47; Indels 35; Gaps 8;

QY 35 YATGTAQAVSEPIDVESHGSGTTPAAGAGSGDDIGYAIWIK--DQVNDVKLVIRNAE 92
Db 98 FSTGT-----PVTVDVDFISLQKAAPQGS--FGFAFEPIKSIDKIDKTVRLTLKH-- 147
QY 93 QLKPY-----FKYLOIQTSGYETNSTALGNFSETRKAVISLNDPSAVIVLDKEDIAV 144
Db 148 ---PVSALLESALSLVAASIVSRADYKESAFG-----SNPVT--GPFKVESYERGQAV 198
QY 145 LYFDKTYNTSIWVGEPDKIIVY 169
Db 199 LVP-----NTYWRQGEDGKALPY 217

RESULT 6
Q8CX23 PRELIMINARY; PRT; 255 AA.
ID Q8CX23
AC Q8CX23;
DT 01-WAR-2003 (TREMELrel. 23, Created)
DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-WAR-2003 (TREMELrel. 23, Last annotation update)
DE cAMP factor.
GN GBS2000.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Meadek T., Zouane M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing

invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766855; CAD47659.1; -.
DR Sagalinst; gbs2000; -.
KW Complete proteome.
SQ SEQUENCE 255 AA; 28374 MW; 266DBB9D0B73562C CRC64;

Query Match 8.7%; Score 91; DB 16; Length 255;
Best Local Similarity 24.5%; Pred. No. 12;
Matches 52; Conservative 31; Mismatches 85; Indels 44; Gaps 10;

QY 5 TLAIAGIIASAALALLAGFATTQSPPLNSFYATGTAQAVSEPIDVES--HLGSITPAAGAQ 63
Db 12 TLVAGALLFSPAVLEHADQVTTTPQVNVHNSNNQAQMAQKLDQDSIQLRNKID--NVQ 69
QY 64 GSD---DIGYAIWIKQVNDVKLVIRL-----NAEQLPKPYKYLQIQT 106
Db 70 GTDYKTYNEAITSVE-----KLKTSLRANPETVYDLSIGSRVEALTDVIE--AITFS 121
QY 107 SGYETNSTALGN---FSETRKAVISLNDPSAVI-----VLDKEDIAVLPDKTGYT 153
Db 122 TQHLANKVSOANIDMGFGITKLVIRILDPFASVDSIKAQVNDVKALEQKLVLPDLKPTD 181
QY 154 NTSIWVGEPDKIIVYNE--TKPVAILNFKAF 183
Db 182 RATIYTKSKLDKEI-WNTRFTRDKKVLNVKEF 212

RESULT 7
Q8CX01 PRELIMINARY; PRT; 255 AA.
ID Q8CX01
AC Q8CX01;
DT 01-WAR-2003 (TREMELrel. 23, Created)
DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-WAR-2003 (TREMELrel. 23, Last annotation update)
DE cAMP factor.
GN CFB OR SAG2043.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AF014283; AAN00902.1; -.
DR TIGR; SAG2043; -.
KW Complete proteome.
SQ SEQUENCE 255 AA; 28374 MW; 266DBB9D0B73562C CRC64;

Query Match 8.7%; Score 91; DB 16; Length 255;
Best Local Similarity 24.5%; Pred. No. 12;
Matches 52; Conservative 31; Mismatches 85; Indels 44; Gaps 10;

QY 5 TLAIAGIIASAALALLAGFATTQSPPLNSFYATGTAQAVSEPIDVES--HLGSITPAAGAQ 63
Db 12 TLVAGALLFSPAVLEHADQVTTTPQVNVHNSNNQAQMAQKLDQDSIQLRNKID--NVQ 69
QY 64 GSD---DIGYAIWIKQVNDVKLVIRL-----NAEQLPKPYKYLQIQT 106
Db 70 GTDYKTYNEAITSVE-----KLKTSLRANPETVYDLSIGSRVEALTDVIE--AITFS 121

```

```

QY 107 SGYETNSTALGN-----FSTKAVISLNDPSAVI-----VLDKEDTAVLYPKTGYT 153
Db 122 TQHLANKVQANIDMGFGITKLVIRILDFFASVDSIKAQVNDVKALEQKVLTYPLDKPTD 181
QY 154 NTSIWVPGPDKIIIVNE--TKPVAILNFKAF 183
Db 182 RATIYTKSLDKKEI-WNTRFTRDKKVLNKEF 212

RESULT 8
Q9YCP2 PRELIMINARY; PRT; 533 AA.
AC Q9YCP2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein APE1216.
GN APE1216.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine Y., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Furahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000061; BA380205.1; -
DR PIR; G72593; G72593.
DR HSSP; P04268; LIC2.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR002592; Rec_signal.
DR InterPro; IPR001865; Ribosomal_S2.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF01664; Rec_signal; 1.
DR Pfam; PF00435; spectrin; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR SMART; SM00150; SPEC; 1.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 533 AA; 56481 MW; E8B53D791C353E31 CRC64;

Query Match 8.7%; Score 91; DB 17; Length 533;
Best Local Similarity 22.4%; Pred. No. 34;
Matches 49; Conservative 24; Mismatches 76; Indels 70; Gaps 9;

QY 5 TLTAAGIITASAAALLAGFATTQSPNLSFYATGTAQVSEPIDVESHLGSIIPAAGAQG 64
Db 17 TLTAAPLAGGAGAGAPSAFVVGSP-----TPVDIVVWLGGEF----- 57

QY 65 SDDIGYAI--VWIKQVNDVKLVTRNA-----EOLKPYFKYIQITSGYET 111
Db 58 -----YDVVEVWLSRGVSEVLIATYLGSGEGVELAGPRLSPGVTLEFRALSGL-- 110

QY 112 NSTALGNSETKAVISLNDPSAVI--VLDKE-----DIAVLVYDK 149
Db 111 --TASAVVEVBEATITLDKQGVGVEVLTAEVGGGGVSVAYQLRMGGVTVAVYDD 168

QY 150 TGYTNTSIWVGPEDPKI-----IVYNETKPVAILNFKAFY 184

us-09-997-807-2.rspt

Db 169 SGRAVASFTVPLPSPGVHVELVYT---PPLWLRFSIY 204

RESULT 9
Q8AV58 PRELIMINARY; PRT; 2169 AA.
AC Q8AV58;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Sidekick-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22217368; PubMed=12230981;
RA Yamagata M., Weiner J.A., Sanes J.R.;
RT "Sidekicks: synaptic adhesion molecules that promote lamina-specific
RT connectivity in the retina.";
RL Cell 110:649-660(2002).
DR EMBL; AF537107; AAN15075.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004896; F:hematopoietin/interferon-class (D200-domain. . .; IEA.
DR InterPro; IPR000282; Cyto receptor_2.
DR InterPro; IPR003962; FhIII_subd.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003599; IG_
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 13.
DR Pfam; PF00047; IG; 6.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 13.
DR SMART; SM00409; IG; 7.
DR SMART; SM00408; IGC2; 6.
DR PROSITE; PSS0835; IG_LIKE; 5.
DR POSITE; 2169 AA; 239477 MW; 06BFD900AEF19BD7 CRC64;

Query Match 8.7%; Score 91; DB 13; Length 2169;
Best Local Similarity 21.6%; Pred. No. 2.5e+02;
Matches 48; Conservative 34; Mismatches 74; Indels 66; Gaps 9;

QY 2 KYTTLAAGIITASAAALLAGFATTQSPNLSFYATGTAQVSEPI----- 47
Db 1376 KFTTVEVGSTVROFTATDL-----TPESAVIFRTSAKTQGWGEPLAATVITTEKRPA 1430

QY 48 -----DVESHLSITPAAGAGSDDIGVAIVMKQVN-----DVKLKVTLR 89
Db 1431 PPQOLITPQSDVSRSLQHLWVFGSDGSSPIRFTVQVRELPGNDWGTYSSTISHEATSC 1490

QY 90 NAEOLKPYFKY-LQIQITSGYETNSTALGNFS-ETKAVISLND-----PSAVIVLDKEDI 142
Db 1491 IIESLNFTSYKLRVK-----ATNDIGSDSYSAETEAVTTLQDVPDPSPSVLV----- 1539

QY 143 AVLYPDKTGYTNTSIWVGPEDKIIIVYNETKPVAILNFKAFY 184
Db 1540 -----TPHTTSSVLVQWQPPKA-----ESLNGLLGYRIY 1570

RESULT 10
Q9I118 PRELIMINARY; PRT; 174 AA.
AC Q9I118;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein PA2464.
GN PA2464.
```

```

OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Slover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzer L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.I., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004674; AAG05852.1; -.
DR PIR; B83337; B83337.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 174 AA; 17626 MW; A00DFC525FCCE2F CRC64;

Query Match      8.6%; Score 90; DB 16; Length 174;
Best Local Similarity 26.0%; Pred. No. 8.6;
Matches 45; Conservative 27; Mismatches 63; Indels 38; Gaps 8;

QY 9 AGIIASAAALALLAGFATTQSPINSFYATGTAQAVSEPIDV-ESHLSGIT-----PA 59
Db Q9C2C4; PRELIMINARY; PRT; 896 AA.
3 ARIAALGLLLPLAAGSAVAEEGSSVKDAKA-AVSSAIEGTGKLLGGVSEGITSGRSA 61
QY 60 AGAAG-----SDDIGYAIWIKQVNDVKLVKTLRNAEQLKPYFKYLQIQTTSGETNSTA 115
Db QGADGAKVSDNAGFA-----EIQGVEVLKTEPDQQLAVTLGFKNDGDA 106

QY 116 ---LGNFSETKAVISLNDP---SAVILKDEDIAVLPDKTGYTNTSIWVPG 161
Db TVRLINLTQTGALVIDNDGYSNALVALANDDDTV--PAKAGIQTFVFEFG 157

RESULT 11
Q9C2C4
ID Q9C2C4 PRELIMINARY; PRT; 896 AA.
AC Q9C2C4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable beta-glucosidase 1.
GN Bin2.40.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hobeisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL513444; CAC28685.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR002772; Glyco_hydro_3C.
DR InterPro; IPR001764; Glyco_hydro_3N.
DR InterPro; IPR000719; Prot_Kinase.
DR Pfam; PF00933; Glyco_hydro_3; 1.
DR Pfam; PF01915; Glyco_hydro_3_C; 1.
DR PRINTS; PR00133; GLHYDRLASE3.

DR PROSITE; PS00775; GLYCOSYL HYDROL F3; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
SQ SEQUENCE 896 AA; 96788 MW; A89848B5DAE8D76B CRC64;

Query Match      8.6%; Score 90; DB 3; Length 896;
Best Local Similarity 21.6%; Pred. No. 87;
Matches 40; Conservative 33; Mismatches 56; Indels 56; Gaps 7;

QY 22 AGFA-----TTQSPINSFYATGTAQAVSEPIDVESHLSGIT----- 57
Db 709 AGFSKVRVRIYWLKSKDADKAYAVGTSS-----SKSGQTYPYEGYSSVKPGV 760
QY 58 PAAGAQSDDIGYAIWIKQVNDVKLVKTLRNA-----EQLKPYFKYLQIQTTSGETNS 113
Db 761 PAGGGGGNPAUFDTI-----LELDVTQVTSRHKGKASQVAYIQFTDSGYDTPI 812
QY 114 TALGNFSETKAVISLNDPNSAVILKDEDIAV-----LYPKTGYTNTS---TWVPG 161
Db 813 IQLRDFAKTKELGTGESETVTLRLRRKDLSDVMDTKQNVVAPGALGANGSKRYIIVLGE 872
QY 162 EPDKI 166
Db 873 GSKDL 877

RESULT 12
Q7WYN3
ID Q7WYN3 PRELIMINARY; PRT; 942 AA.
AC Q7WYN3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulosomal scaffoldin adaptor protein B.
GN SCAB.
OS Acetivibrio cellulolyticus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Acetivibrio.
OX NCBI_TaxID=35830;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12867464;
RA Xu Q., Gao W., Ding S.-Y., Kenig R., Shoham Y., Bayer E.A., Lamed R.;
RT "The Cellulosome System of Acetivibrio cellulolyticus Includes a Novel
RT Type of Adaptor Protein and a Cell Surface Anchoring Protein.";
RL J. Bacteriol. 185:4548-4557(2003).
DR EMBL; AY221112; AAP48995.1; -.
SQ SEQUENCE 942 AA; 100358 MW; CABB5A638B352CA9 CRC64;

Query Match      8.6%; Score 90; DB 2; Length 942;
Best Local Similarity 24.2%; Pred. No. 93;
Matches 53; Conservative 29; Mismatches 101; Indels 36; Gaps 10;

QY 10 GITASAAALALLAGFATTQSPINSFYATGTAQAVS-EPIDVESHLSGITPAAGAQSDDI 68
Db 113 GILNFSAKAYSLDDYRTAAAP-----EQTGVAVVVKFKVLEETSSISFEDTTSVPNAIDG 168
QY 69 GYAIWIKQVND--VKLVKTLRNAEQLKPYFKYLQIQTTSGETNSTALGN-FSETKAV 125
Db 169 TVLEFDWNGDRILQSGSYVIQPAVINLDMIKASY-----ITMGYDKNAEYGEIITATVKI 222
QY 126 ISLNDPNSAVILKDEDIAVLP--DKTG--YTNFSIWVPGEPDKIIVYNETKPV----- 175
Db 223 NKITNFGYGVNIKYDPTVLQAVNPKTGVAYTNTSSLSPTSGE---LLVSEYDGPVQGVHK 279
QY 176 ---AIIKFKAFYEAKGMLFDSLIP-----VIFNQVLQ 205
Db 280 ISEGILNLSRYSYTALEYVRASESEPETGLAVVGVFKVLQ 318

RESULT 13
Q9S0W9
ID Q9S0W9 PRELIMINARY; PRT; 449 AA.
AC Q9S0W9;

```

DT	01-MAY-2000 (T-EMBLrel. 13, Created)	DR	InterPro; IPR004903; SLAP.
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)	DR	Pfam; PF03217; SLAP; 1.
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)	DR	PRINTS; PRO1729; SURFACELAYER.
DE	Proteinase.	SQ	SEQUENCE 450 AA; 47826 MW; D38D15A958B2C13 CRC64;
GN	Lactobacillus helveticus.		
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;		
OC	Lactobacillus.		
OX	NCBI_TaxID=1587;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CP790;		
RA	Yamamoto N., Shinoda T., Takano T.;		
RT	"Molecular cloning and sequence analysis of a gene encoding an		
RT	extracellular proteinase from Lactobacillus helveticus CP790.";		
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB026985; BAA86287.1; -		
DR	PIR; JC7306; JC7306.		
DR	GO; GO:0009274; C:cell wall (sensu Bacteria); IEA.		
DR	GO; GO:0030115; C:S-layer; IEA.		
DR	GO; GO:0005199; F:structural constituent of cell wall; IEA.		
DR	InterPro; IPR004903; SLAP.		
DR	Pfam; PF03217; SLAP; 1.		
DR	PRINTS; PRO1729; SURFACELAYER.		
SQ	SEQUENCE 449 AA; 47798 MW; 8F2A22A2E5077326 CRC64;		
Query Match 8.6%; Score 89.5; DB 2; Length 449;			
Best Local Similarity 24.0%; Pred. No. 36;			
Matches 52; Conservative 33; Mismatches 87; Indels 45; Gaps 9;			
QY	11 IIASAAALALLAGFATTQSPNSFYATGTAQAVSEPIDVESHLS-----ITPA-- 59	QY	11 IIASAAALALLAGFATTQSPNSFYATGTAQAVSEPIDVESHLS-----ITPA-- 59
Db	8 VSAAMAAALAVPAATAMPVNAATVTTSTTNKPTVDLSGAGSVSESKDTVNVTPSFT 67	Db	8 VSAAMAAALAVPAATAMPVNAATVTTSTTNKPTVDLSGAGSVSESKDTVNVTPSFT 67
QY	60 -----AGAGQSDDIGYAIWIKQVNDVKLVTLRNAEQPKYF-----KYLQ 102	QY	60 -----AGAGQSDDIGYAIWIKQVNDVKLVTLRNAEQPKYF-----KYLQ 102
Db	68 LLSAAKGIPATLQGSIEASLNGTSVTADVADVAKDVTLDGKGVAVSYDKNTLTNKLSD 127	Db	68 LLSAAKGIPATLQGSIEASLNGTSVTADVADVAKDVTLDGKGVAVSYDKNTLTNKLSD 127
QY	103 IQITSGYETNSTALGNSETKA-----VISLNPASVIV-----LDKEDIAVLYPKDTGYTN 154	QY	103 IQITSGYETNSTALGNSETKA-----VISLNPASVIV-----LDKEDIAVLYPKDTGYTN 154
Db	128 VKAGDDYTMVLSGVG-FSGKANAGKTLTFLKPEGVTEGANYNKDHHKVTL-DQYGNVS 185	Db	128 VKAGDDYTMVLSGVG-FSGKANAGKTLTFLKPEGVTEGANYNKDHHKVTL-DQYGNVS 185
QY	155 TSIWVPEPDKIIIVNETKPVAILNFKAFYEAKGML 191	QY	155 TSIWVPEPDKIIIVNETKPVAILNFKAFYEAKGML 191
Db	186 GLKFV---ISKVAYDSANTNAV-----SFYDAKSLV 215	Db	186 GLKFV---ISKVAYDSANTNAV-----SFYDAKSLV 215
RESULT 15			
Q60045	PRELIMINARY; PRT; 1148 AA.		
ID	Q60045;		
AC	Q60045;		
DT	01-NOV-1996 (T-EMBLrel. 01, Created)		
DT	01-NOV-1996 (T-EMBLrel. 01, Last sequence update)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)		
DE	Polygalacturonase precursor (EC 3.2.1.82).		
GN	PGLA.		
OS	Thermoanaerobacter thermosulfurogenes (Clostridium		
OS	thermosulfurogenes).		
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;		
OC	Thermoanaerobacteriaceae; Thermoanaerobacterium.		
OX	NCBI_TaxID=33950;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=EM1;		
RA	Matuschek M., Sahm K., Bahl H.;		
RL	Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES		
CC	(POLYGALACTURONASES).		
CC	-1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.		
DR	EMBL; U50951; AAB08040.1; -		
DR	PIR; S72635; S72635.		
DR	GO; GO:0005618; C:cell wall; IEA.		
DR	GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.		
DR	GO; GO:0004650; F:polygalacturonase activity; IEA.		
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.		
DR	InterPro; IPR003961; FN_III.		
DR	InterPro; IPR008957; FN_III-like.		
DR	InterPro; IPR000743; Glyco_hydro_28.		
DR	InterPro; IPR001119; SLH.		
DR	Pfam; PF00041; fn3; 2.		
DR	Pfam; PF00295; Glyco_hydro_28; 1.		
DR	Pfam; PF00395; SLH; 3.		
DR	SMART; SMO0060; FN3; 2.		
DR	PROSITE; PS00502; POLYGALACTURONASE; 1.		
DR	PROSITE; PS01072; SLH DOMAIN; 2.		
KW	Cell wall; Glycosidase; Hydrolase; Signal.		
FT	SIGNAL 1 30 POTENTIAL		
FT	CHAIN 31 1148 POLYGALACTURONASE		
SQ	SEQUENCE 1148 AA; 122642 MW; E95HD0F2529FAL0E CRC64;		
Query Match 8.5%; Score 88.5; DB 2; Length 1148;			
Best Local Similarity 19.1%; Pred. No. 1.6e+02;			
Matches 37; Conservative 38; Mismatches 82; Indels 37; Gaps 6;			

QY 20 LLAGFATTCSPINSPYATCTAQAUSEPIDVESHLSITPAAGAGQSDDDIGYALVWIK-DQ 78  
Db 14 LLAFLMFTMPMSKAPADTTSSGPAPVDLQVLSSVT-----DTGFTLWVHKPDN 64  
QY 79 VNDV-KLKVTLRMAEOLKPYFKYLQIITSGY-----ETNSTALGNF--SE 121  
Db 65 YSDITDYKITVSDSVYEQVYASENOTVASQYIKQFYDNNVGDLDKDDNGNTWMSAYKISM 124  
QY 122 TXAVISLDNPSAVIVLDKEDIAVLXPDKTGYTNTSIWVPGBDKIIIVYNETKPVAILNFK 181  
Db 125 HSFVLTGLKENTLYTIQVQSDANKNTSTPVITITQSTAPSTPSE-----NIINVE 174  
QY 182 AFYEAKGMLFDLSL 195  
Db 175 STGAVGDGVILADDV 188

Search completed: April 6, 2004, 12:41:16  
Job time : 47 secs